

TARTU ÜLIKOOL  
LOODUS- JA TÄPPISTEADUSTE VALDKOND  
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT  
BIOINFORMAATIKA ÕPPETOOL

**Geneetiliselt muundatud taimede tuvastamine sekveneerimise  
toorlugemitest kasutades kindla pikkusega  $k$ -meere**

Magistritöö

30 EAP

Siimo Kangruoja

Juhendaja

M.Sc Kairi Raime

TARTU 2020

# INFOLEHT

## **Geneetiliselt muundatud taimede tuvastamine sekveneerimise toorlugemitest kasutades kindla pikkusega $k$ -meere**

### **Lühikokkuvõte:**

Põllumajanduse tootlikkuse suurendamiseks on kasutusele võetud biotehnoloogilised meetodid, mis võimaldavad luua uute tunnustega sorte geneetilise muundamise teel. Selleks, et tagada tarbijale valikuvabadus ning tõkestada autoriseerimata GM sortide levikut toidu ja sööda ahelas on vaja täpseid ja suure läbilaskevõimega tuvastusmeetodeid. Praegused lahendused pole võimalised efektiivselt identifitseerima pidevalt suurenevat sortide hulka. Käesolevas töös eesmärk on testida  $k$ -meeride kasutamise võimalikust transgeense sordi tuvastamiseks sekveneerimise toorlugemitest. Antud meetodit saaks rakendada GM sortide tuvastamiseks toiduainete või sööda toorainest või töödeldud valmis produktidest.

**Märksõnad:**  $k$ -meer, GMO, genoomsed markerid

**CERCS kood:** B110 – Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

## **Detection of modified organisms from raw sequencing data using $k$ -mers of specific length**

### **Abstract:**

To increase agricultural yield focus has turned to biotechnological methods, which can create varieties with new traits. To enable consumers the freedom of choice and stop the spread of unauthorized genetically modified crops in the food chain proper control is needed. Currently, used methods are not able to handle the expanding amount of enhanced strains. This study aims to test the possibility of using  $k$ -mers to detect GM crop event-specific markers of transgenic origin and to identify them in raw sequence reads. The given method could be used to detect GM crops in the food or feed supply chain.

**Keywords:**  $k$ -mers, GMO, genomic markers

**CERCS:** B110 Bioinformatics, medical informatics, biomathematics, biometrics.

# SISUKORD

SISUKORD .....	3
KASUTATUD LÜHENDID .....	5
1. KIRJANDUSE ÜLEVAADE .....	7
1.1. GM taimed ja nende konstrueerimine .....	7
1.2. GM taimede tuvastamise olulisus .....	9
1.3. GMO tuvastamise meetodid .....	10
1.4. K-meeride põhine lähenemine .....	14
2. EKSPERIMENTAALNE OSA .....	16
2.1. Töö eesmärgid .....	16
2.2. Materjal ja metoodika .....	16
2.2.1. Töös kasutatud andmed .....	16
2.2.2. Töövoo loomine GM sordi-spetsiifiliste <i>k</i> -meeride leidmiseks .....	16
2.2.3. Maisi ja soja GM sordi-spetsiifiliste <i>k</i> -meeride leidmine .....	17
2.2.3.1. Positiivne kontroll .....	18
2.2.3.2. Negatiivne kontroll .....	19
2.2.4. Maisi ja soja liigispetsiifiliste <i>k</i> -meeride tuvastamine ja testimine .....	19
2.3. Tulemused .....	21
2.3.1. Töövoo GM sündmuses esinevate sordi-spetsiifiliste <i>k</i> -meeride leidmiseks ...	21
2.3.2. GM maisi ja soja sordi-spetsiifilised <i>k</i> -meerid .....	22
2.3.3. Positiivne kontroll .....	24
2.3.4. GM sordi-spetsiifilised <i>k</i> -meerid muundamata genoomides .....	26
2.3.5. Soja ja maisi liigispetsiifilised <i>k</i> -meerid .....	27
2.3.6. GM sündmuse tuvastamine .....	28
2.4. ARUTELU .....	31
KOKKUVÕTE .....	34
SUMMARY .....	35

TÄNUAVALDUSED.....	36
KASUTATUD KIRJANDUSE LOETELU .....	37
KASUTATUD VEEBIAADRESSID .....	44
LISA .....	45
LIHTLITSENTS.....	51

## KASUTATUD LÜHENDID

CryAb/c - *Bacillus thuringiensis*-est pärit protoksiinikristallid Ab/C

GM – geneetiliselt muundatud

GMDD - GMO detektsiooni meetodite andmebaas

GMO – geneetiliselt muundatud organism

k-meer – oligomeer pikkusega „k“

p35S – Lillkapsa mosaiikviiruse 35S promootor

qPCR – kvantitatiivne polümeraasi ahelreaktsioon

t35S – Lillkapsa mosaiikviiruse 35S terminaator

T-DNA – transfeer DNA

Ti-plasmiid – tuumori induktsiooni võimega plasmiid

tNOS – *Agrobacterium tumefaciens* nopaliini süntaasi terminaator

vip3A – *Bacillus thuringiensis*-est pärit vegetatiivne insektitsiidne valk 3A (*vegetative insecticidal protein 3A*)

## SISSEJUHATUS

Maailma aina suureneva elanikkonna toitmiseks on vaja suurendada põllumajanduslikku tootlikust. Seda aga pole võimalik teha läbi põllumaa suurendamise. Üha enam on kasutatud biotehnoloogilisi meetodeid, mis võimaldavad läbi geneetilise modifitseerimise luua taimesorte uute tunnustega ning tõsta seeläbi taimede saagikust. Regulatsioonid geneetiliselt muundatud põllukultuuridele varieeruvad oluliselt üle maailma erinevates riikides. Ühes riigis lubatud ja laialt levinud muundatud põllukultuuri võivad paratamatult sattuda ka teise riiki, kus need on keelatud.

Selleks, et tagada tarbijatele valikuvabadus ning tõkestada autoriseerimata GMO-de levikut toiduahelas on oluline GM rutiinne testimine. Praegu kasutatakse selleks peamiselt PCR-põhiseid meetodeid. Antud meetodid on ennast tõestanud, kuid ei tule toime aina suureneva GM sortide arvuga. Teise põlvkonna sekveneerimismeetodite ja joondamisvaba *k*-meeride põhise lähenemise kombineerimine pakuks uudset lähenemist GM sündmuste tuvastamiseks otse sekveneerimise toorlugemitest. Seni pole GM taimede tuvastamiseks *k*-meeride põhiseid meetodeid rakendatud.

Käesoleva magistritöö kirjanduse ülevaates kirjeldatakse GM põllukultuuride ajalugu, nende tegemise mehhanisme ja seni enim kasutatud detektsiooni meetodeid. Praktilise osa eesmärk on töötada välja töövoog sordi- ja sündmuse-spetsiifiliste *k*-meeride leidmiseks ning testida *k*-meeride kasutamise võimalikkust GM sündmuste tuvastamiseks etteantud proovi sekveneerimise lugemitest, soja ja maisi näitel.

# 1. KIRJANDUSE ÜLEVAADE

## 1.1.GM taimed ja nende konstrueerimine

GMO ehk geneetiliselt muundatud organism on organism, välja arvatud inimene, mille geneetilist materjali (DNA-d) on muudetud viisil, mida ei toimu looduslikult läbi ristamise ja/või rekombinatsiooni (DIREKTIIV 2001/18/EL).

Transgeensete taimesortide loomise põhjuseks oli soov pääseda mööda traditsioonilise sordiaretuse piirangutest. Antud lähenemine võimaldab sortidesse viia uusi tunnuseid, mida looduses sellel liigil või lähedastel liikidel ei eksisteeri. Uuteks tunnusteks võivad olla resistentsus teatud kahjurputukate ning haigustekitajate vastu, herbitsiidi tolerantsus, põuataluvus, suurenenud toiteväärtus jne. Esimene GM (geneetiliselt muundatud) taim loodi 1983. aastal (Fraley et al., 1983). Esimese geneetiliselt muundatud põllukultuuriga jõuti turule 1992. aastal (James, 1997).

GM põllukultuure kasvatati 2015. aastal 179,9 miljonil hektaril, sellest 116 780 hektarit oli Euroopas kasvatatud GM mais (James, 2015). Enim kasvatatakse muundatud taimi Ühendriikides, kus 94% sojaoast (*Glycine max*) ja 93% kasvatatavast maisist (*Zea mays*) on geneetiliselt muundatud (Perry et al., 2016).

GM taimede loomise saab (tinglikult) jaotada kuueks etapiks. Esmalt tuleb leida kasuliku tunnusega organism. Seejärel antud tunnust kodeeriv geen tuvastatakse ja eraldatakse. Pärast seda toimub geeni disainimine. Järgnevalt viiakse tunnust kodeeriv geen taime genoomi ning valitakse välja taimed, mille genoomi õnnestus geen viia (Boyle, 2011). Viimasena ristatakse väljavahitud taimed tagasi muundamata taimedega (Gao et al., 2013).

Muutmata transgeenil on uude organismi viies madal või ettearvamatu ekspressioonitase. See tuleneb erinevast koodonkasutusest ning RNA regulatsioonist fülogeneetiliselt kaugetes organismides (Novoa et al., 2012; Streatfield, 2007). Seetõttu on geeni disainimine oluline ekspressiooni taseme tõstmiseks. Selleks tehakse erinevaid muudatusi: muudetakse koodonkasutus peremeesorganismi sarnaseks, muudetakse geenide GC sisaldus muundamata organismi oma sarnaseks, välditakse ebastabiilsete RNA motiivide kasutamist ja eemaldatakse potentsiaalsed RNA sekundaarstruktuur translatsiooni alguskoha ligidalt (Jackson et al., 2014). Joonisel 1 on kujutatud 1800 nt pikkune *Bacillus thuringensise* ssp. kurstaki CryA(b) (protoksiinikristallid A(b)) geeni muundamata järjestuse ja maisi kolme optimeeritud transgeeni järjestuse vaheline joondus (Holst-Jensen et al., 2012).

	80	90	100	110	120	130	140
B. thuringiensis kurstaki	..... ..... ..... ..... ..... ..... ..... .....	ATTCTTTATAATTGTTTAAGTAACCCCTGAAGTAGAAGTATTAGGTGGAGAAAGAAATAGAAACTGGTTACA					
Bt11_CryIA	..... ..... ..... ..... ..... ..... ..... .....	ATTCCTATCAACTGCTTGAGTAACCCAGAAGTTGAAGTACTTGGTGGAGAACGCATTGAAACCGGTTACA					
Putative Bt176 CryIA	..... ..... ..... ..... ..... ..... ..... .....	ATCCCTACAACCTGCCTGAGCAACCCCGAGGTGGAGGTGCTGGGCGGCGAGCGCATCGAGACCGGCTACA					
Mon810_CryIA	..... ..... ..... ..... ..... ..... ..... .....	ATCCCGTACAACCTGCCTCAGCAACCCCTGAGGTGAGGTGCTCGGCGGTGAGCGCATCGAGACCGGTTACA					
	150	160	170	180	190	200	210
B. thuringiensis kurstaki	..... ..... ..... ..... ..... ..... ..... .....	CCCCAATCGATATTTCTTGTGCGCTAACGCAATTTCTTTGAGTGAATTTGTTCCCGGTGCTGGATTTGT					
Bt11_CryIA	..... ..... ..... ..... ..... ..... ..... .....	CTCCCATCGACATCTCTTGTCTTGACACAGTTTCTGCTCAGCGAGTTCTGTGCCAGGTGCTGGGTTCTGT					
Putative Bt176 CryIA	..... ..... ..... ..... ..... ..... ..... .....	CCCCCATCGACATCAGCCTGAGCCTGACCCAGTTCTGCTGAGCGAGTTCTGTGCCCGGCGCCGGCTTCGT					
Mon810_CryIA	..... ..... ..... ..... ..... ..... ..... .....	CCCCCATCGACATCTCCCTCTCCCTCAGCAGTTCTGCTCAGCGAGTTCTGTGCCAGGCGCTGGCTTCGT					
	220	230	240	250	260	270	280
B. thuringiensis kurstaki	..... ..... ..... ..... ..... ..... ..... .....	GTTAGGACTAGTTGATATAATATGGGGATTTTGGTCCCTCTCAATGGGACGCAATTTCTTGTACAAAT					
Bt11_CryIA	..... ..... ..... ..... ..... ..... ..... .....	TCTCGGACTAGTTGACATCATCTGGGGTATCTTTGGTCCATCTCAATGGGATGCATTCTTGGTGCAAAAT					
Putative Bt176 CryIA	..... ..... ..... ..... ..... ..... ..... .....	GCTGGGCGTGGTGACATCATCTGGGGCATCTTCCGCCCCAGCCAGTGGGACGCTTCTTGGTGCAATC					
Mon810_CryIA	..... ..... ..... ..... ..... ..... ..... .....	CCTGGGCGTGGTGACATCATCTGGGGCATCTTGGGCCCTCCAGTGGGACGCTTCTTGGTGCAATC					

**Joonis 1.** *Bacillus thuringiensis* ssp. kurstaki geeni cryA(b) ja kolme geneetiliselt muundatud maisi (Bt11, Bt176, MON810) optimeeritud järjestuse osaline joondus (Holst-Jensen et al., 2012).

Peale optimeerimist sisestatakse transgeen muundatava taime genoomi ning selleks on erinevaid meetodeid: bioballistika, *Agrobacterium*'i vahendatud transformatsioon, viirused, elektroporatsioon, genoomi editeerimine jne. Eelnimetatud lähenemistest kasutatakse enim bioballistikat ja *Agrobacterium*'i vahendatud transformatsiooni.

Bioballistika on meetod, mille puhul tulistatakse taimerakke DNA-ga kaetud kulla või volframi partiklitega. Taimerakku jõudes eraldub DNA osakese küljest ning integreerub tuuma. Antud meetodil on olulisi puuduseid: insertsioonide arv on kõrge ning insertsioonid paiknevad tihti tandeemselt.

Enim kasutatakse transgeensete taimede loomiseks mullabakteri *Agrobacterium tumefaciens*i vahendatud transformatsiooni. *Agrobacterium* on võimeline üle kandma oma Ti-plasmiidi (tuumori induktsiooni võimega plasmiid) T-DNA (Transfeer DNA) regiooni taime genoomi. *Agrobacterium*'i T-DNA piirkond on mõlemalt poolt piiritletud 25 nukleotiidi pikkuse mitteidentse kordusjärjestusega. T-DNA protsessimine saab alguse parempoolsest kordusjärjestusest ja lõpeb tavaliselt vasakul kordusjärjestusel. Mõnikord ei lõpetata vasakul kordusjärjestusel protsessimist ja taime genoomi jõuab lisaks T-DNA järjestusele ka osa vasakpoolsest külgnest järjestusest (Hwang et al., 2017).

2016. aastal käidi välja hüpotees, mille järgi viib *Agrobacterium*'i T-DNA integreerimist läbi DNA polümeraas  $\theta$  taime homolog TEICHI. Antud polümeraas teostab vigaderohket mikrohomoloogiate poolt suunatud kaheahelaliste katkete parandamist. TEICHI ühendab T-DNA taime genoomis eelnevalt eksisteerivasse kaheahelalisse katkesse ning sünteesib ühenduskohta nn „filler DNA“, mille pikkus varieerub 1 – 51 bp, mis omab enamasti



mikrohomoloogiat taime genoomis lähedal asuva alaga (Kregten et al., 2016; Windels, 2003). Niinimetatud „*filler DNA*“ on kasulik tuvastamiseks, sest loob integreerumisel unikaalse järjestuse mõlemasse transgeense kasseti otsa.

Taimedes on olemas mehhanism homoloogiliseks rekombinatsiooniks, kuid selle toimumissagedus somaatilistes rakkudes on  $8.4 \times 10^5$  kuni  $2.2 \times 10^6$  korda väiksem mittehomoloogilisest rekombinatsioonist (Wright et al., 2005). *Agrobacteriumi* infektsioon ei tõsta homoloogilise rekombinatsiooni tekkesagedust, seetõttu integreerub T-DNA genoomis juhuslikku kohta (Kim ja Gelvin, 2007). T-DNA integreerumise kohta plastiidide genoomi on liiga vähe informatsiooni, et võiks selle tõenäosust hinnata.

## **1.2.GM taimede tuvastamise olulisus**

GMO-de tuvastamine toidust on oluline, et tagada tarbijale valikuvabadus (Dong et al., 2008), saada ülevaade GMO-de liikumisest terves toiduahelas (Fraiture et al., 2015) ning takistada autoriseerimata GMO-de levikut (S. R. M. Broeders et al., 2012). GMO-de kasvatamisele ja kasutamisele kehtivad seadused erinevad riigiti oluliselt. Osades riikides (Argentiina, Kanada, USA jne) ei ole vajalik GM organisme sisaldavate toidu ja sööda märgistamine, teistes (Austraalia, Hiina, EU jne) on see kohustuslik. Massiprotsent, millest alates tuleb GMO sisaldav toode märgistada varieerub oluliselt, olles 0% kuni 5%. Euroopa Liidus on piiriks kehtestatud kuni 0,9% komponendi kogu massist (Kamle ja Ali, 2013). Näiteks, kui toode sisaldab 50% maisi, siis GM maisi võib olla 0,45% kogu tootest. Antud piir on kehtestatud, kuna teatud protsessides on tehniliselt võimatu erinevatest saadetistest/kaubalaadungitest pärineva toormaterjali segunemise vältimine. Antud piirmäär ei kehti, kui GMO on lisatud tahtlikult. Autoriseerimata GMO-sid toodetes olla ei tohi. (Weighardt, 2010)

GMO „saastuse“ liiga hiline avastamine võib omada suurt mõju majandusele ning inimeste usaldusele. 2000. aastal saastus USA-s mitte-GMO mais autoriseerimata GMO maisiga ning jõudis seeläbi inimeste toidulauale. Antud juhtum vähendas märgatavalt maisi eksporti välisriikidesse ning USA maisi hind langes 6,8% (Carter ja Smith, 2007). 2006. aastal saastusid USAs muundamata riisi põllud autoriseerimata GM riisiga. Saastunud riisi impordi tõkestamiseks Euroopasse hakkas EL oluliselt rohkem testima USA-st tulevaid riisisaadetisi. USA kasvatajatel saamata jäänud tulu on hinnatud kuni 400 miljoni dollarini (Thompson et al., 2015). Tõsiseks probleemiks on Hiinast saabuval autoriseerimata GMO-dega saastunud riisi saadetised. Selle tõkestamiseks võeti 2012. aasta alguses vastu määrus (Direktiiv 2011/884/EU), mis nõuab Hiinast tulevate riisi saadetiste süstemaatilist kontrolli GM markerite vastu.

Euroopa liidus on suur nõudlus loomasööda järgi, mille täitmiseks imporditi 2018. aastal 14 miljonit tonni soja ning 19 miljonit tonni maisi ([http://europa.eu/rapid/press-release\\_IP-19-161\\_en.htm](http://europa.eu/rapid/press-release_IP-19-161_en.htm)) (kasutatud 25.07.2018). 2016. aastal oli 85% imporditud sojast ning 5 - 25% maisist GMO (SWD 8.3.2016 VER 61).

Vaatamata suurele impordile, kasvatatakse muundatud põllusorte Euroopas ainult Hispaanias (124 227 hektarit) ja Portugalis (7 308 hektarit). Kultiveeritakse ainult ühte muundatud põllukultuuri, maisi sort MON810. Antud sordile on lisatud *Cry* geen, mis kodeerib protoksiinikristalle. Antud produkt omab insektsiidseid omadusi, tekitades putukate seedetrakti auke. Maailmas kasvatatakse enim GM maisi ja GM soja sorte, millele on antud herbitsiidi tolerantsuse ja/või putukaresistentsed omadused (ISAAA, 2017).

### **1.3.GMO tuvastamise meetodid**

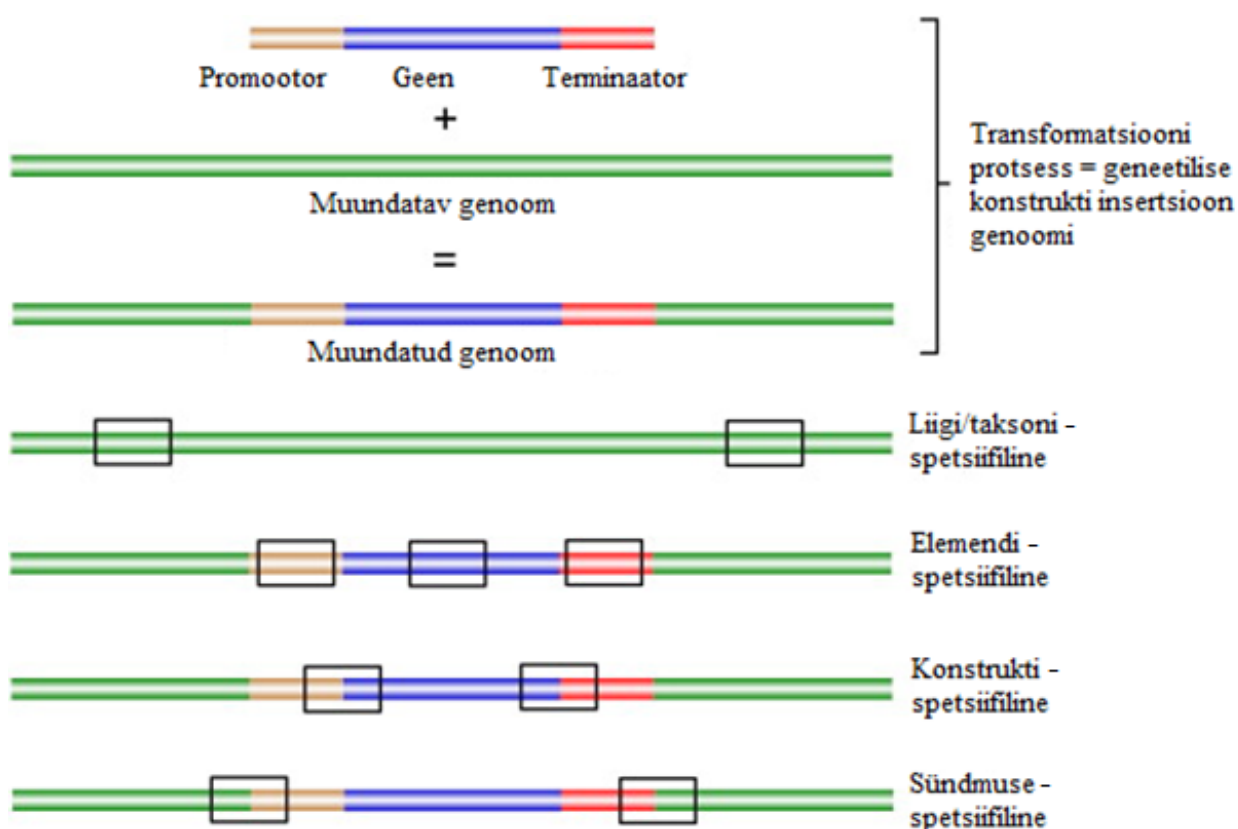
GMO tuvastamise meetodid võib jagada kaheks - kaudsed (valgupõhised) ja otsesed (DNA-põhised) (Fraiture et al., 2015).

Võõrvalkude detekteerimiseks kasutatakse peamiselt kahte meetodit, näiteks ribatest (*strip test*) või ELISA (ensüümi-immuunanalüüs). Valgupõhiste meetodite kasutamine on odav ja kiire ning nende rakendus ei nõua keerukat väljaõpet (Mazzara et al., 2013). Tõsiseks puuduseks on uute antikehade arendamisele kuluv aeg (Mazzara et al., 2013). Erinevalt DNA-põhistest meetoditest ei saa valgupõhiste lähenemistega tuvastada regulatoorsete elementide olemasolu (promootorid, enhanserid ja terminaatorid), seetõttu puudub teisest liigist juhusliku saaste detektsiooni võimalus (Mazzara et al., 2013). Valgupõhiste meetodite puuduseks on ka asjaolu, et valgu ekspressioonitase varieerub erinevates taime kudedes ja arengustaadiumites ning valgud lagunevad või denatureerivad toidu valmistamise protsessides (Grothaus et al., 2006).

DNA-põhised meetodid võimaldavad tuvastada regulatoorseid elemente ning on seetõttu oluliselt rohkem kasutatud. DNA-põhiseid meetodeid on mitmeid: PCR põhised (PCR, qPCR, dPCR, PCR CGE, DNA walking), LAMP, mikrokiip analüüsid ning sekveneerimise-põhised (Sanger ja NGS (teise põlvkonna sekveneerimine)) (Fraiture et al., 2015).

GM taimede tuvastamise saab spetsiifilisuse järgi jagada neljaks erinevaks tasemeks (joonis 2):

- liik/takson – määratakse geneetiliselt muundatud taimeliigi või taksoni olemasolu proovis
- element – tuvastatakse promootori, enhanseri, geeni või terminaatori olemasolu
- konstrukti – määratakse kindlate elementide kõrvuti paiknemine
- sündmus – tuvastatakse GMO kasseti integratsioonis tekkinud unikaalset järjestust (Holst-Jensen et al., 2012).



**Joonis 2.** Transformatsiooni protsess ning neli erinevat tuvastuse taset (liigi/taksoni-, elemendi-, konstrukti- ning sündmuse-spetsiifiline tuvastus) (Holst-Jensen et al., 2012, muudetud).

DNA-põhistest meetoditest enim kasutatud meetod on PCR (Angers-Loustau et al., 2014). Lühikest amplokoni suurust kasutades (üldjuhul umbes 100 bp) saab antud meetodiga GMO-d tuvastada nii töötlemata kui ka väga töödeldud toiduainetest (S. Broeders et al., 2012). PCR-põhise meetodi puudus on see, et meetod võimaldab tuvastada ainult elemente, konstrukte ja sündmusi, millele on välja arendatud praimerid.

GMO-de detekteerimiseks kasutatakse kolme sammu. Esmalt testitakse võimaliku GMO olemasolu kasutades levinumaid elemente nagu p35S (Lillkapsa mosaiigiviiruse 35S promootor), tNOS (nopaliini süntaasi terminaator *Agrobacterium Tumefaciens*'ist). Järgmisena kasutatakse konstrukti tasemel tuvastamist ja/või sündmuse tasemel tuvastamist. Viimasena tuvastatakse liigi olemasolu ning vajadusel sündmus (S. Broeders et al., 2012). Kui eelkirjeldatud lähenemisega sündmust identifitseerida ei suudeta, siis järeldatakse, et on tegemist autoriseerimata GMO-ga. Antud lähenemisel PCR-põhiste meetoditega on mitu puudust, esmalt nõuab see PCR mitmekordset tegemist, mis on töömahukas. Teisalt tuvastatud autoriseeritud GMO võib varjata sarnaste elementidega autoriseerimata GMO-d. Kolmandaks, elemendi tuvastamine ei tõesta koheselt GMO olemasolu, vaid võib olla proovi sattunud saastuse kaudu (Fraiture, Herman, De Loose, et al., 2017).

Euroopas kasutusloa saanud GM sündmusi on hetkel 43, kuuest erinevast liigist, ning lisaks 9 on veel taotlusjärgus (EUROPEAN COMMISSION, 2017). USAs on kasutamisloa saanud 204 erinevat muundamise sündmust 20 liigist ("GM Crop Events approved in United States of America," 2019.). On välja töötatud qPCR meetodeid, millega saab samaaegselt tuvastada 39 erinevat GM sündmust seitsmest erinevast liigist (Querci et al., 2009).

qPCR läbilaskvuse suurendamiseks on arendatud *multiplex* strateegia, mis kujutab endast mitme PCR reaktsiooni läbiviimist samas reaktsioonikambris (Chamberlain et al., 1988). Viie kuni kümne samaaegse PCR reaktsiooni puhul suureneb oluliselt süsteemi keerukus ning hakkab esinema arvestataval hulgal valepositiivseid tulemusi. Lisaks on PCR meetodiga keeruline minna rutiinselt kasutatavalt 96 auguga plaadilt suurema reaktsioonide arvu peale, sest see nõuaks lisainvesteeringuid automatiseerimisse, tööprotsessi ümberkujundamisse ning töötajate väljaõppesse (Querci et al., 2010).

Toidu kuumtöötlemine põhjustab DNA fragmenteerumist ning raskendab seeläbi GMO-de PCR-põhist tuvastamist, antud raskuse ületamiseks kasutatakse lühikese amplikoniga PCR-i, kuid antud lähenemine ei võimalda nii hästi eristada sarnaseid järjestusi (Little, 2014). Toidu mitmekülgne kooslus muudab raskeks praimerite koguste optimeerimise (King et al., 2008; Markoulatos et al., 2002). Lisaks on PCR meetodil vajadus paralleelselt sooritada positiivne kontroll, see nõuab aga referentsmaterjali, mis ei ole alati kättesaadav (Holst-Jensen et al., 2012). Eelmainitud puuduste taustal vajab qPCR-i jätkusuutlikus GMO peamise tuvastusmeetodina analüüsimist (Querci et al., 2010).

Lisaks PCR meetoditele, on üha rohkem hakatud GMO tuvastamiseks kasutama ka sekveneerimispõhiseid meetodeid (Košir et al., 2018), mis erinevalt PCR põhistest meetoditest võimaldab tuvastada oluliselt rohkem muundamise sündmuseid. Sekveneerimispõhise

lähenemise saab jagada kaheks: suunatud ning ülegenoomne sekveneerimine. Sangeri sekveneerimist kasutatakse tüüpiliselt oma väikese läbilaskevõime tõttu suunatud sekveneerimisega (Fraiture et al., 2015).

Juhul, kui proovis tuvastatakse teadaolevaid transgeenseid elemente, kasutatakse järgnevas etapis suunatud sekveneerimist, mille puhul suurendatakse uuritava DNA osakaalu enne järjestusanalüüsi teostamist. Sellega saab vähendada identifitseerimiseks vajaliku sekveneerimise mahtu. Üks suunatud sekveneerimise meetoditest on *DNA walking*, mis võimaldab amplifitseerida teadaoleva järjestuse kõrval asuvat tundmatut järjestust. Hetkel on välja arendatud *DNA walkingul* põhinevad meetodid pCambia (Fraiture et al., 2014), t35S (Lillkapsa mosaiikviiruse 35S terminaator), p35S ja tNOS (*Agrobacterium tumefaciens* nopaliini süntaasi terminaator) (Fraiture et al., 2015), CryAb/c (*Bacillus thuringiensis*-est pärinevad protoksiinikristallid Ab/c) (Fraiture et al., 2018) ning vip3A (*Bacillus thuringiensis*-e vegetatiivne insektitsiidne valk 3A) (Liang et al., 2014) elementide tuvastamiseks. Antud elementidega on võimalik detekteerida suurt osa teadaolevatest GM taimedest. Peale uuritava järjestuse amplifitseerimist, amplikonid sekveneeritakse kasutades teise põlvkonna sekveneerimismeetodit. Lugemite analüüsiks kasutatakse tihti joondamist ning andmebaasidest vastete otsimist (Fraiture et al., 2017). Joondamine võimaldab tuvastada ka seni kirjeldamata sorte, kuid on väga aja- ja töömahukas ning nõuab bioinformaatilisi oskusi. Eelnevalt kirjeldamata GMO-de tuvastamisel joondatakse lugemid peremeesorganismi (näiteks soja või maisi) referentsjärjestuste vastu ning lugemid, mida ei õnnestunud osaliselt või täielikult joondada analüüsitakse edasi.

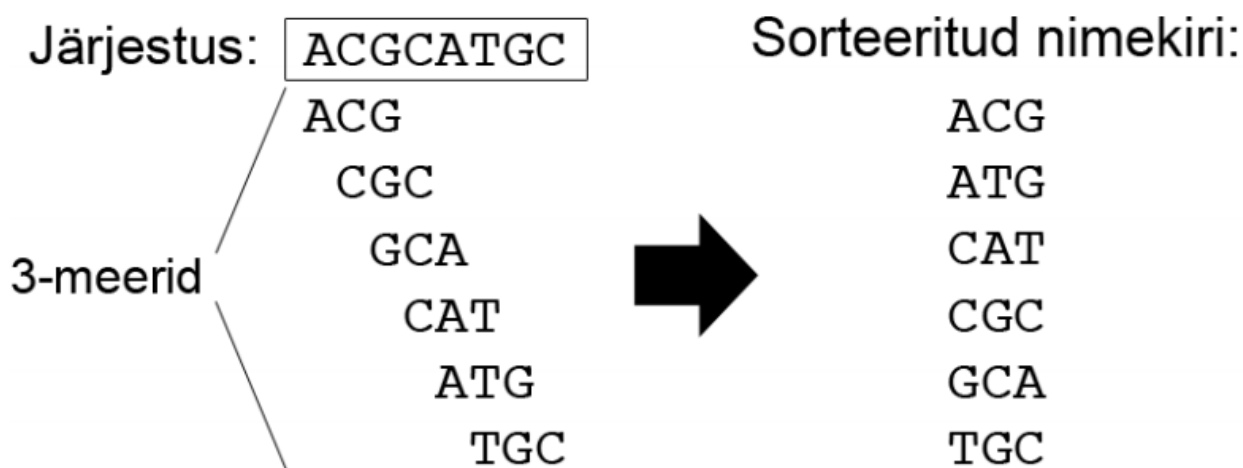
Üha populaarsemaks on muutunud ülegenoomne sekveneerimise põhised tuvastamise meetodid, mille puhul enne sekveneerimist spetsiifilist DNA regiooni ei amplifitseerita. Antud lähenemist kasutatakse regulaarselt metagenoomide analüüsiks, kuid GMO-de tuvastamisel on antud lähenemist rakendatud harva (Wetterstrand, 2018). Sekveneerimisel saadud toorandmete analüüsiks kasutatakse tavaliselt joondamist või *k*-meeride lugemist. Esimese puhul joondatakse lugemid referentsgenoomidele (Ripp et al., 2014). GMO tuvastamisel on seni rakendatud ainult joondamist (Arulandhu et al., 2018). Antud protsessis otsitakse sekveneerimisel saadud lugemitele andmebaasist vasteid, määrates seeläbi proovis esinevad liigid. Töömahu vähendamiseks limiteeritakse referents-genoomide arvu, vaatamata sellele on antud meetod ressursi- ja ajamahukas. Teise puhul loetakse kokku proovis esinenud *k*-meerid ja nende sagedused ning tuvastatakse selle abil proovis esinenud liigid (Raime ja Remm, 2018; Roosaare et al., 2017; Wood ja Salzberg, 2014).

## 1.4.K-meeride põhine lähenemine

K-meerid on kõik erinevad kindla pikkusega „ $k$ “ oligomeerid, mis sisalduvad mingis genoomijärjestuses (joonis 3). Võimalik  $k$ -meeride arv järjestuse kohta saadakse kasutades valemit

$$L - k + 1$$

kus „ $L$ “ on järjestuse pikkus ning „ $k$ “ on  $k$ -meeri pikkus (Altschul et al., 1990). Tänapäeval levinud bioinformaatika tööriistadest kasutavad  $k$ -meeride põhist lähenemist BLAST (Altschul et al., 1990), genoomi assambleerijaid, näiteks SPAdes (Bankevich et al., 2012), ja metagenoomi analüüsijad, millega tuvastatakse proovis sisalduvaid liike ja tüvesid (Raime ja Remm, 2018; Rizk et al., 2013; Roosaare et al., 2017; Wood ja Salzberg, 2014).



**Joonis 3.** Järjestus, mis on tehtud  $k$ -meerideks ( $k = 3$ ) ning sorteeritud (Vaher, 2016).

Metagenoomidest bakterite tüvede tuvastamist  $k$ -meeride abil võimaldab näiteks programm Strainseeker (Roosaare et al., 2017). Selle kasutamiseks on vaja assambleeritud bakteritüvede kõrge kvaliteediga genoome ning juhtpuud, mis kirjeldaks fülogeneetilist suhet erinevate tüvede vahel (Roosaare et al., 2017). Metagenoomi analüüsimises on  $k$ -meeride lugemine oluliselt kiirem kui joondamine (Lindgreen et al., 2016) ning ei nõua nii suurt arvutusvõimsust, kuna ei säilitata informatsiooni  $k$ -meeride paiknemise kohta, vaid loetakse ainult nende esinemissagedust. See võimaldab leida vasteid oluliselt kiiremini kui joondamispõhine BLAST (Wood ja Salzberg, 2014).

K-meeride põhise tuvastamise eeliseid saaks üle kanda ka GMO detekteerimisse. Toidu kuumtöötlemine põhjustab DNA fragmenteerumise tõttu GMO tuvastamisel arvestaval hulgal

valenegatiivseid tulemusi (Mano et al., 2017). Erinevalt teistest meetoditest ei oma  $k$ -meeride põhisele tuvastamisele olulist mõju sekveneerimislugemi pikkus, kuna otsitava järjestuse pikkus saab olla väga lühike (kuni 32 nukleotiidi).

## 2. EKSPERIMENTAALNE OSA

### 2.1. Töö eesmärgid

Käesoleva töö eesmärk on töötada välja töövoog GM sordi-spetsiifiliste *k*-meeride leidmiseks, leida GM sortide- ja sündmuse-spetsiifilised *k*-meerid ning liigispetsiifilised *k*-meerid GM maisi ja soja tuvastamiseks täisgenoomi sekveneerimise lugemitest ning lõpuks testida *k*-meeride spetsiifilisust ja tuvastamise tundlikkust muundamata ning muundatud taimede täisgenoomi sekveneerimise toorlugemite abil.

### 2.2. Materjal ja meetoodika

#### 2.2.1. Töös kasutatud andmed

Töös kasutatud GM taimesortide nimekirja aluseks võeti Euroopa GMO referents andmebaas (<http://www.euginius.eu/> 20.03.2019). Antud andmebaasis oli kokku 703 sissekannet erinevate geneetiliselt muundatud organismide kohta. Geneetiliselt muundatud taimesorte oli neist 693. See koosnes omakorda 303 unikaalsest muundamise sündmusest, ülejäänud 390 oli saadud unikaalsete GM taimede ristamisel. GMO järjestustele puudub ühtne andmebaas, järjestuste leidmiseks kasutati erinevaid andmebaase, nagu Google Scholar (<https://scholar.google.com/>), lens.org (<https://www.lens.org>), NCBI (NCBI Resource Coordinators, 2017), GMO JRC amplicon (<https://gmo-crl.jrc.ec.europa.eu/jrcgmoamplicons/>)(Angers-Loustau et al., 2014) ning GMDD (GMO detektsiooni meetodite andmebaas) ([gmdd.sjtu.edu.cn](http://gmdd.sjtu.edu.cn), Dong et al., 2008). Eelnevalt loendatud andmebaasidest oli saadaval järjestused 88 GMO sündmusele 11 liigist (lisa 1), mis võimaldavad sündmuse-spetsiifilist tuvastamist. Maisi oli nendest 17 ning soja 11 sorti. Kõiki maisi 31 ja soja 17 leitud GM insertsiooni järjestusi kasutati *k*-meeride nimekirjade koostamisel.

#### 2.2.2. Töövoo loomine GM sordi-spetsiifiliste *k*-meeride leidmiseks

GM sordi-spetsiifiliste *k*-meeride leidmise esimeses etapis on vaja eemaldada *k*-meerid, mis pole GMO-le spetsiifilised. Andmebaasidest kogutud GM järjestused sisaldavad enamasti ka taime genoomset päritolu flankeeruvaid järjestusi. Need võivad töö hilisemas osas anda valepositiivseid tulemusi, kuna sisaldavad sordi-spetsiifilisi SNP, mille välja sõelumine pärast *k*-meeride nimekirja koostamist on raskendatud. Selle vältimiseks koostati programmeerimiskeeles Pythoni 3.6 skript, mis eemaldas antud järjestustest 5' ja 3' otstes olevad genoomsed järjestused. Genoomi ja transgeeni vaheline piir saadi järjestuse allikast või



järjestuse võrdlemisel NCBI Nucleotide (<https://www.ncbi.nlm.nih.gov/nucleotide>) ja NCBI Assembly andmebaasis (<https://www.ncbi.nlm.nih.gov/assembly>) olevate DNA järjestusi kasutades NCBI BLAST vaikeseadetega.

GM sordi järjestustest sordi-spetsiifiliste *k*-meeride nimekirjade koostamiseks kasutati programmipaketti GenomeTester4 (Kaplinski et al., 2015). Antud pakett sisaldab programmi glistmaker võimaldab luua *k*-meeri nimekirju toorlugemistest ja assambleeritud järjestustest, *k*-meeri pikkusega 1 kuni 32 nt.

Ainult GM päritolu *k*-meeridest koosneva nimekirja loomiseks eemaldati eelnevalt loodud nimekirjadest *k*-meerid, mis esinesid doonor-organismides (organismides, kust pärinevad transgeensed elemendid), peremeesorganismis (antud taimeliigi või -sordi muundamata genoomis) või teistes GM sortides. Doonor-organismides esinevate *k*-meeride eemaldamine on oluline, kuna mõned organismid on levinud taimepatogeeneid (näiteks lillkapsa mosaiikviirus), esinevad rikkalikult mullas (näiteks *Agrobacterium tumefaciens*) või kasutatakse toiduks ja/või söödaks (näiteks mais, soja, mis on lisaks ka peremeesorganismid). Peremeesorganismi, soja ja maisi, genoomis esinevate *k*-meeride eemaldamine on oluline, et välistada juhuslikult identse *k*-meeri olemasolu, mis võiks viia valepositiivse tulemuseni juhul, kui analüüsitavas proovis on ka vastava liigi mitte-GM taime DNA-d. Seeläbi saadud nimekirjad kirjeldavad muundamisspetsiifiliste *k*-meeri põhiste markerite arvu üle terve inserteerunud kasseti. GM päritolu *k*-meeride nimekirjast eemaldati teiste GM sortide genoomijärjestustes esinevad *k*-meerid, et loodud nimekirjad kirjeldaksid paremini ainult spetsiifilises GM sordis asuvaid *k*-meere.

### **2.2.3. Maisi ja soja GM sordi-spetsiifiliste *k*-meeride leidmine**

Väljatöötatud töövoogu GM sordi/sündmuse spetsiifiliste *k*-meeride leidmiseks rakendati, et leida kahe enamlevinud GM taimeliigi (maisi ja soja) GM sordi-spetsiifilisi *k*-meere.

Antud töö fookuses olnud GM soja ja maisi muundamisel kasutatud doonor-organismide liigi nimekiri pärineb Eugenis (<https://www.euginius.eu>) andmebaasist. Doonor- ning peremeesorganismide genoomijärjestused pärinevad NCBI Assembly andmebaasist. Juhul, kui doonor-organismi täisgenoom polnud saadaval, kasutati SRA toorlugemi andmeid (*Andropogon gerardii*, *Avena sativa*, *Saccharum officinarum* ja *Discosoma sp*). Täielik nimekiri eemaldamiseks kasutatud järjestustest on toodud lisa 2.

Kahel doonor-organismil polnud NCBI andmebaasides saadaval järjestusi: hübriidpetuunia (*Petunia hybrida*) ja julia priimula (*Primula juliae*). Mitte ükski neist pole söögitaim, mistõttu nende sattumine toidu või sööda sisse pole tõenäoline.

Töös eemaldati maisi ja sojaga samasse sugukonda kuuluvate söödavate liikide referentsgenoomides esinevad *k*-meerid, et välistada elemendi või sündmuse-spetsiifiliste *k*-meeride esinemist lähedaste liikide muundamata genoomides. Mais kuulub kõrreliste (*Poaceae*) sugukonda, NCBI andmebaasis oli täisgenoomijärjestus olemas antud taksoni viiel söödaval liigil, mille järjestusi analüüsil kasutati: itaalia kukeleib (*Setaria Italica*) (GCA\_000263155.2), *Cenchrus americanus* (GCA\_002174835.2), looduslik kaheteranisu (*Triticum dicoccoides*) (GCA\_002162155.2), harilik hirss (*Panicum miliaceum*) (GCA\_002895445.2) ja pikateraline riis (*Oryza sativa*) (GCA\_001618795.1). Soja kuulub liblikõieliste (*Fabaceae*) sugukonda. NCBI andmebaasis on täisgenoom olemas ja analüüsil kasutati järgmiste sama sugukonna söödavate liikide genoomijärjestusi: harilik kikerhernes (*Cicer arietinum*) (GCA\_000331145.1), harilik tuvihernes (*Cajanus cajan*) (GCA\_000340665.1), mungvigna (*Vigna radiata*) (GCA\_000816755.2), asukivigna (*Vigna angularis*) (GCA\_001190045.1), ahtalehine lupiin (*Lupinus angustifolius*) (GCA\_001865875.1), harilik maapähkel (*Arachis hypogaea*) (GCA\_003086295.1), harilik vigna (*Vigna unguiculata*) (GCA\_004118075.1) ja metsik soja (*Glycine soja*) (GCA\_004193775.2).

Optimaalseima *k*-meeri pikkuse leidmiseks koostati *k*-meeride nimekirjad erinevate pikkustega (16, 18, 20, 22, 24, 26, 28, 30 ja 32 nukleotiidi) ning võrreldi nendes esinenud *k*-meeride arvu ning valiti *k*-meeri pikkus, mis andis enim sordi-spetsiifilisi *k*-meere.

Insertsioonil tekkinud ainulaadsete *k*-meeride leidmiseks tuleb leida *k*-meerid, mis paiknevad kasseti insertseerumisel taime genoomi tekkinud „*filler DNA*“ regioonis, nimetatakse käesolevas töös sündmuse-spetsiifilisteks *k*-meerideks. Antud ala piiritlemiseks kasutati NCBI Blast otsingut, kus otsiti GMO järjestusele vasteid NCBI Nucleotide ning NCBI Assembly andmebaasidest, kasutades vaikeväärtusi. GM sündmuse tuvastamiseks ei kasutatud vektorjärjestusi, kuna sama GM kassetti on võimalik kasutada mitme GM taimesordi loomisel (Näiteks mais BT10 ja mais BT11).

#### **2.2.3.1. Positiivne kontroll**

Andmebaasides pole kättesaadavaid GM taimede assambleeritud genoome ega toorlugemi andmeid (välja arvatud GM papaia). Selleks et kontrollida kas leitud GM soja ja maisi spetsiifilised *k*-meerid on tuvastatavad muundatud taimede täisgenoomi sekveneerimise lugemitest ning kui suur hulk lugemeid selleks vaja on, oli vajalik töö fookuses olnud maisi ja soja muundatud sortide täisgenoomi sekveneerimise toorlugemid luua simulatsiooni teel. Insertsiooni koha leidmiseks kasutati transgeense kasseti (päripärisid [lens.org](http://lens.org) andmebaasist)

kõrval olevaid genoomseid alasid, millele otsiti vaste referentsgenoomidest ning NCBI Assembly andmebaasis olevaid maisi (GCA\_000005005.6) ja soja (GCA\_000004515.4) täisgenoomi järjestusi.

Simuleerimiseks kasutati ART simulaatorit (Huang et al., 2012), kuna see võimaldab simuleerida erinevate sekveneerimisplatvormide (Illumina, Roche ja SOLiD) sekveneerimise toorlugemeid. Sätetena valiti Illumina HiSeq 2500 sekvenaator, mis on üks enim maisi ja soja sekveneerimiseks kasutatud platvorm SRA andmebaasi. Lugemi pikkuseks määrati 150, kuna see oli lähim võimalik keskmisele Illumina platvormi paarisotsaliste (*paired-end*) lugemite pikkusele (maisi puhul 247.7030616 ja sojal 218.1806). Simulatsioon tehti kolmes korduses kasutades erinevaid fikseeritud veaprofiile 1554663544, 1221330211 ja 1443552433.

#### **2.2.3.2. Negatiivne kontroll**

Loodud sordi ja sündmuse-spetsiifiliste  $k$ -meeride spetsiifilisuse kontrollimiseks otsiti nendes sisalduvaid  $k$ -meere NCBI SRA andmebaasis olnud erinevate muundamata soja ja maisi proovide täisgenoomi sekveneerimise lugemitest. Testimiseks valiti proovid, mis olid sekveneeritud kasutades Illumina HiSeq 2500 sekvenaatori paarisotsalised lugemite meetodit, ilma eelselekteerimist kindlate regioonide suhtes (*random*) tegemata. Otsingu tulemustest valiti 20 soja toorlugemi järjestust (SRR6796817, SRR11191191, SRR1302621, SRR1585205, SRR2424011, SRR2424012, SRR1533336, SRR4252451, SRR4453817, SRR496796, SRR497630, SRR497634, SRR497635, SRR497636, SRR5480161, SRR5516309, SRR5516310, SRR6796793, SRR7817178 ja SRR8301914), mis sisaldasid kokku 13.7 miljardit lugemit ning koosnesid 928.1 miljardist nukleotiidist. Maisi sortides olnud  $k$ -meere testiti 20 maisi toorlugemi (SRR5826130, SRR7477674, SRR1575500, SRR7477670, SRR7477668, SRR1575518, SRR7477714, SRR6967945, SRR1575503, SRR1575513, SRR6967943, SRR5826129, SRR11302094, SRR7477673, SRR11301678, SRR5725900, SRR7467446, SRR5272164, SRR5272162 ja SRR4039072) järjestusel, mis koosnesid kokku 13.6 miljardist lugemist ning milles oli 1.834 triljonit nukleotiidi.

#### **2.2.4. Maisi ja soja liigispetsiifiliste $k$ -meeride tuvastamine ja testimine**

Maisi ja soja liigispetsiifiliste  $k$ -meeride leidmiseks kasutati PlantTaxSeeker-it (Raime ja Remm, 2018). Antud töövoog võimaldas leida kloroplasti järjestustest taimeliigi spetsiifilisi  $k$ -meere ning tuvastada neid teise põlvkonna sekveneerimise lugemitest (Raime et al., 2020). Kuna hetkel teostatakse PCR testimise käigus ka liigi määramine, tuleks seda samuti teostada ka  $k$ -meeri põhise tuvastamise puhul. Lisaks tulevikus, kui on olemas võimekus elemendi või

konstruktsiooni tasemel tuvastamiseks, on antud informatsioon oluline, et analüüsida uuritavaid GMO-sid.

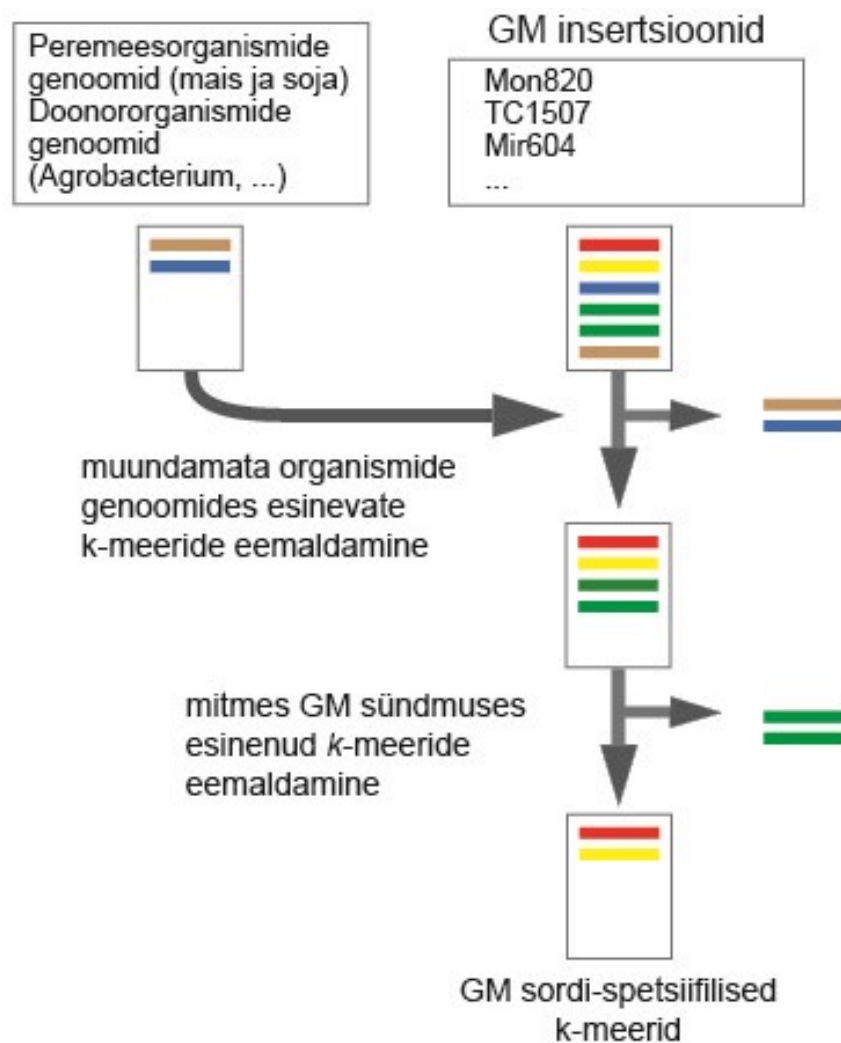
Soja liigispetsiifiliste *k*-meeride nimekirja koostamiseks leiti esmalt kõigis kolmes soja kloroplasti genoomi järjestuses (NCBI Nucleotide andmebaas DQ317523.1, NC\_007942.1 ja CM010429.1) esinevad ühised *k*-meerid, mis ei esinenud üheski teises NCBI Nucleotide (19.06.2019) RefSeq andmebaasis olemasolevas 2976 taime kloroplasti järjestuses. Järgnevalt eemaldati lähedastes liikides (NCBI SRA andmebaas, (*Glycine latifolia*) SRR6234223, (*Glycine dolichocarpa*) SRR1175585, (*Glycine syndetika*) SRR1176843, (*Glycine tomentella*) SRR6823487) ja maisi (NCBI SRA andmebaas SRR1575509) toorlugemites esinenud *k*-meerid. Viimasena testiti allesjäänud *k*-meeride olemasolu kümnel NCBI SRA (Leinonen et al., 2011) andmebaasist alla laetud soja proovil (NCBI SRA andmebaas SRR1302613, SRR1302628, SRR2202261, SRR2424014, SRR5085797, SRR5479850, SRR5479862, SRR5479877, SRR5480106 ja SRR5480145).

Maisi liigispetsiifiliste *k*-meeride nimekirja koostamisel leiti *k*-meerid, mis esinesid üheksast NCBI andmebaasis olevast maisi assambleeritud kloroplasti genoomist (NCBI andmebaas NC\_001666.2, X86563.2, AY928077.1, KP966117.1, KP966116.1, KP966115.1, KP966114.1, KF241981.1 ja KF241980.1,) vähemalt kaheksas ning ei esinenud teiste NCBI nucleotide (19.06.2019) RefSeq piiranguga andmebaasis olemas oleva 2976 taimeliigi kloroplasti genoomides. *k*-meeride nimekirja spetsiifilisuse parandamiseks eemaldati *k*-meerid, mis esinesid lähedaste liikide (NCBI SRA andmebaas (*Hordeum vulgare*) SRR3655668, (*Avena sativa*) SRR6056489, (*Oryza sativa*) SRR3297190, (*Sorghum bicolor*) SRR072231, (*Zea luxurians*) SRR088692, (*Triticum aestivum*) SRR1170587, (*Cenchrus Americanum*) SRR5204425 ja (*Saccharum officinarum*) SRR7771854) ja soja (SRR1302613) täisgenoomi sekveneerimise toorlugemites sagedusega rohkem kui kümme korda. Viimasena testiti alles jäänud *k*-meeride esinemist kümne maisi proovi täisgenoomi sekveneerimise lugemites (NCBI SRA andmebaas SRR5976424, SRR7477706, SRR5976220, SRR065912, SRR7477683, SRR5725906, SRR065920, SRR5152975, SRR5152976, SRR5962677).

## 2.3. Tulemused

### 2.3.1. Töövoog GM sündmuses esinevate sordi-spetsiifiliste $k$ -meeride leidmiseks

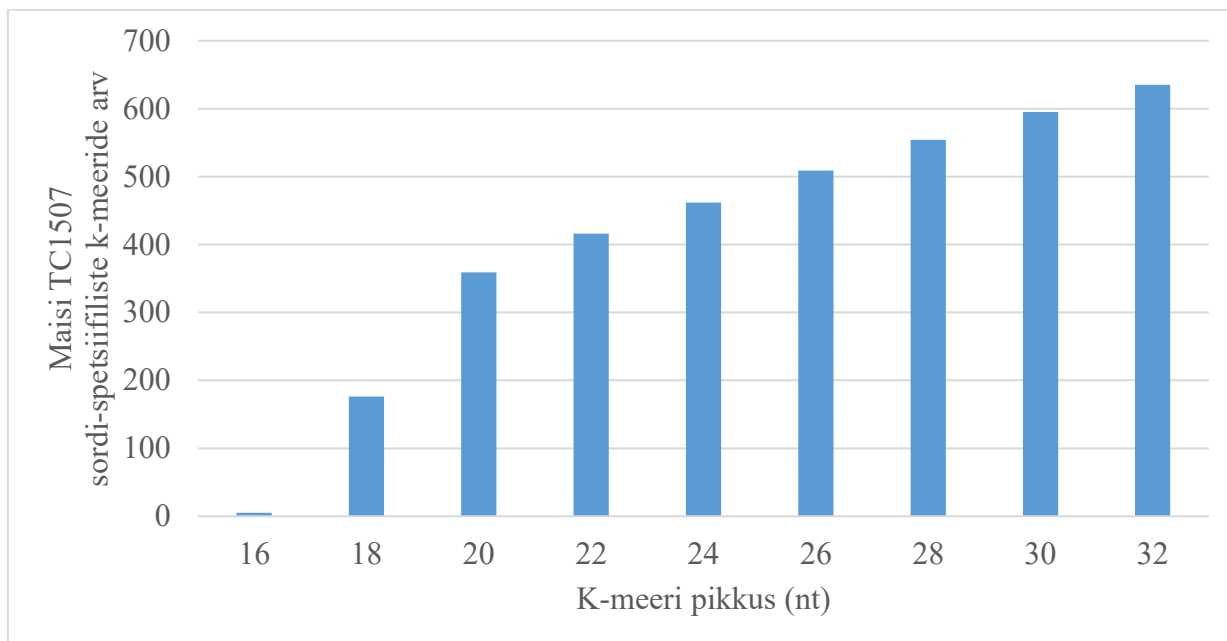
Sorti kirjeldavate  $k$ -meeride leidmiseks koostati Pythoni programmeerimiskeeles töövoog (joonis 4), mis rakendab GenomeTester programmikomplekti programme. Antud töövoogu üheks sisendiks on GM sordi järjestused (GM insertide järjestused, millel inserti ümbritsenud muundamata peremeesorganismi genoomi piirkonnad on eemaldatud) ning teiseks sisendiks peremees- ning doonororganismide järjestused. Esimeses etapis luuakse GM sündmuste järjestustest  $k$ -meeride nimekirjad. Antud nimekirjadest eemaldatakse doonor- ning peremeesorganismide genoomides esinevad  $k$ -meerid. Järgnevalt eemaldatakse mitmes GM sündmuses nimekirjas dubleeritud  $k$ -meerid. Seeläbi saadakse sordi-spetsiifiliste  $k$ -meeride arv iga sordi kohta ning sordi-spetsiifilised  $k$ -meeride nimekirjad (mis sisaldavad  $k$ -meeride järjestusi ja sagedust). Järgnevalt eemaldatakse mitmes GM sündmuse nimekirjas dubleeritud  $k$ -meerid. Seeläbi saadakse sordi-spetsiifilised nimekirjad. Töövoogu skriptid on kättesaadavad käesoleva töö autorilt.



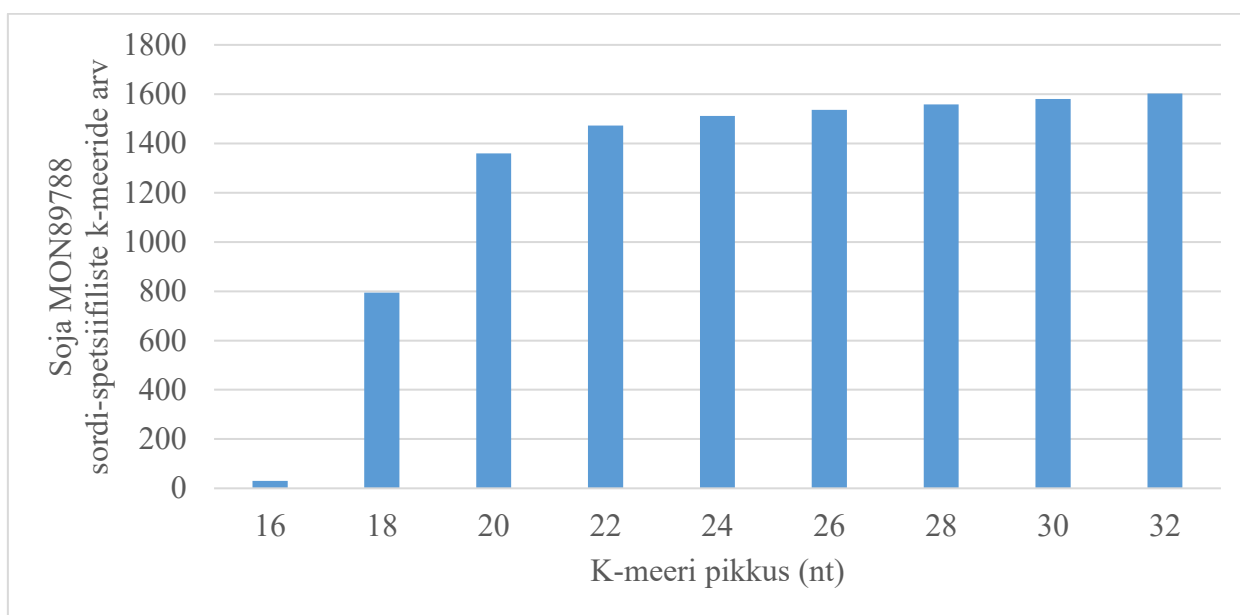
**Joonis 4.** GM sordi-spetsiifiliste *k*-meeride leidmise töövoog ja selle etapid. Esimeses etapis luuakse GM sordi järjestustest *k*-meeride nimekiri, millest eemaldatakse doonor- ning peremeesorganismide genoomides esinevad *k*-meerid. Järgnevalt eemaldatakse *k*-meerid, mis esinevad ka teiste GM sortide puhul. Töövoo tulemusena saadakse GM sordi-spetsiifiliste *k*-meeride nimekiri.

### 2.3.2. GM maisi ja soja sordi-spetsiifilised *k*-meerid

Töö järgmistes osades kasutatakse *k*-meeri pikkust 32 nukleotiidi, mis andis suurimad GM sordi-spetsiifiliste *k*-meeride arvud. Joonisel 5 ja joonisel 6 on välja toodud erinevate *k*-meeride pikkuste puhul tuvastatud GM sordi-spetsiifiliste *k*-meeride arvud GM maisi TC1507 ja soja MON87708 sordi kohta. Tulemused näitavad spetsiifiliste *k*-meeride arvu kasvu *k*-meeri pikkuse kasvades.



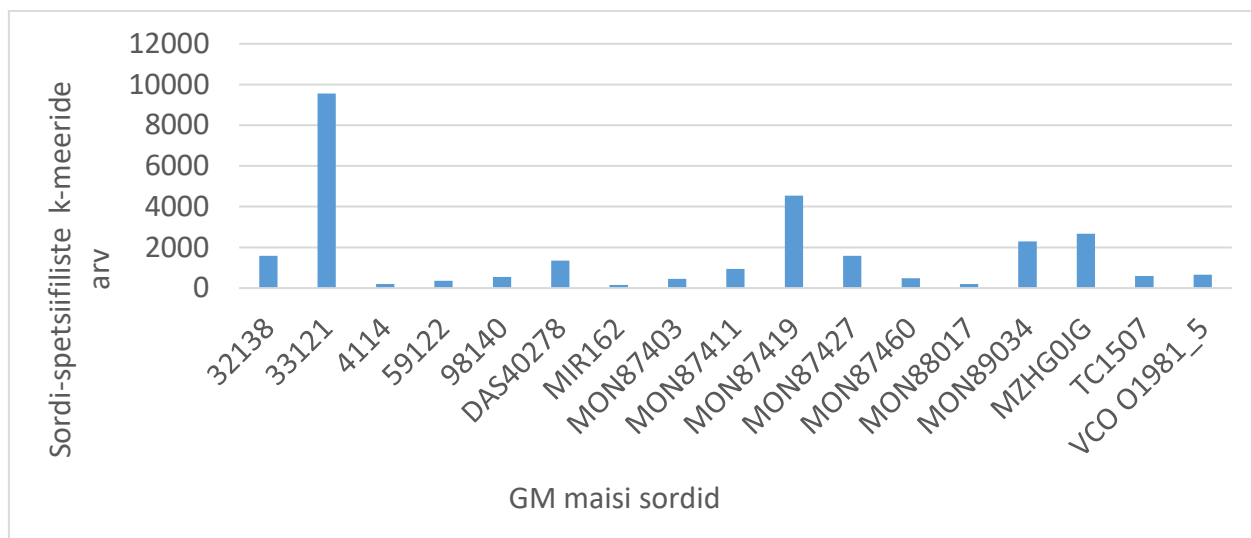
**Joonis 5.** GM muundatud maisi TC1507 sordi-spetsiifiliste  $k$ -meeride arv erinevate  $k$ -meeri pikkuste juures.



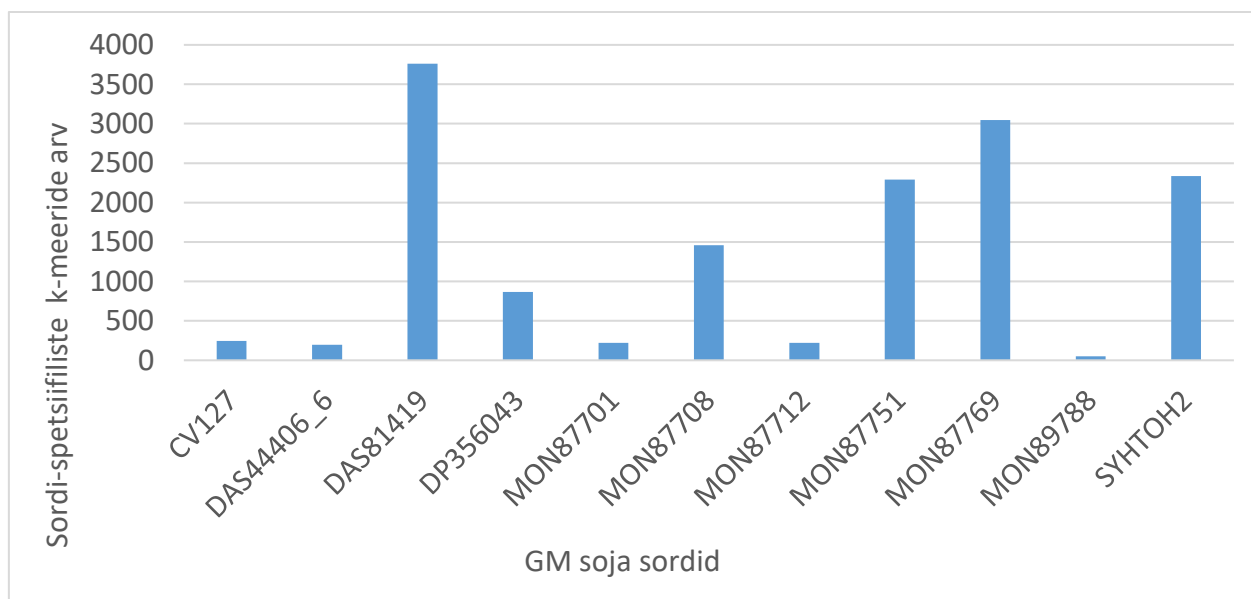
**Joonis 6.** GM muundatud soja MON87708 sordi-spetsiifiliste  $k$ -meeride arv erinevate  $k$ -meeri pikkuste juures.

GM sordi-spetsiifiliste  $k$ -meeride hulk varieerub oluliselt erinevate maisi ja soja sortide vahel, sõltudes insertiooni pikkusest ja spetsiifilisusest (joonis 7 ja joonis 8). Näiteks GM maisi sort 33121 insertiooni suurus on 24 160 nt, sisaldades 9557 sordi-spetsiifilist  $k$ -meeri. Maisi sort

VCO 01981\_5 insertioon on 3 698 nt, milles paiknes 660 sordi-spetsiifilist *k*-meeri. GM soja sort DAS81419 insertiooni suurus on 12 557 nt sisaldades endas 3763 *k*-meeri. Soja sort MON87701 insertiooni suurus on 6 497 nt ning selles leidub 222 sordi-spetsiifilist *k*-meeri.



**Joonis 7.** Tuvastatud sordi-spetsiifiliste *k*-meeride arv erinevatel GM maisi sortidel.



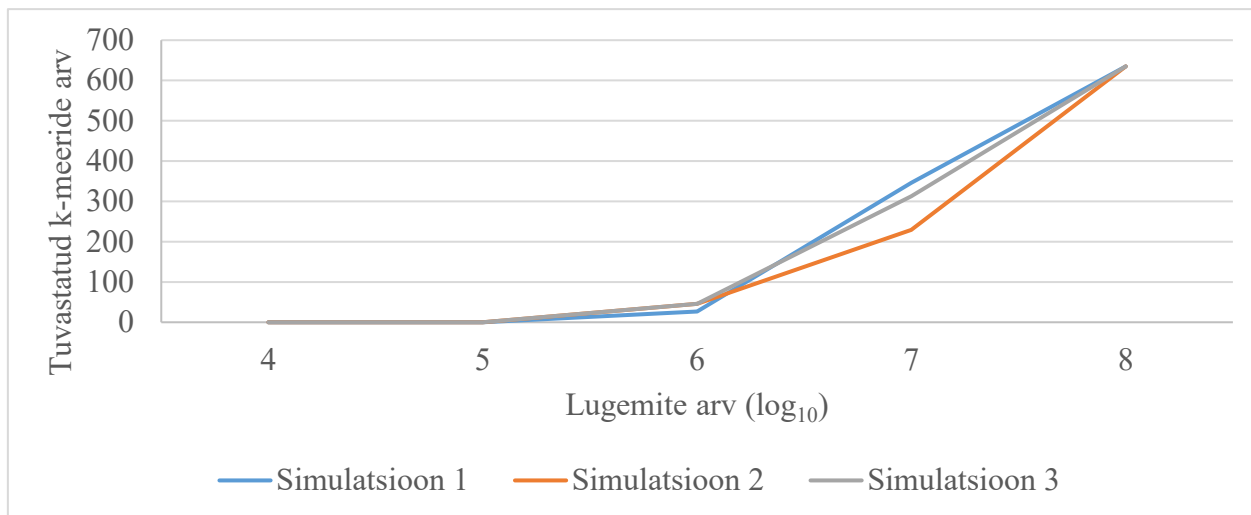
**Joonis 8.** Tuvastatud sordi-spetsiifiliste *k*-meeride arv erinevatel GM soja sortidel.

### 2.3.3. Positiivne kontroll

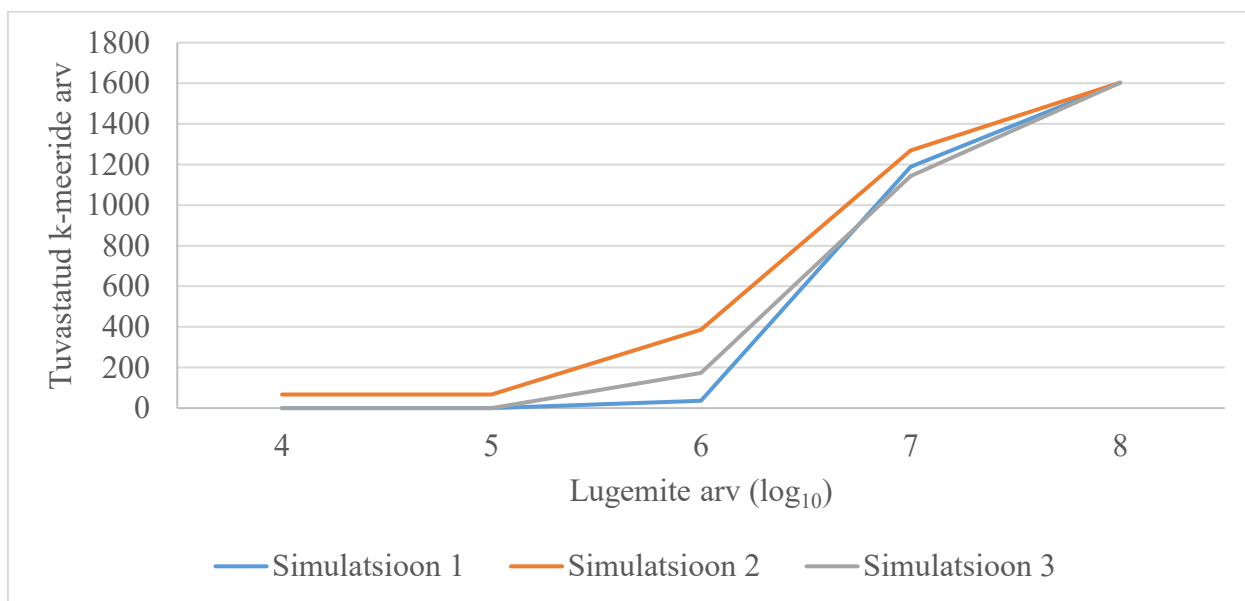
Selleks et hinnata kui palju lugemeid on vaja kõikide *k*-meeride tuvastamiseks täisgenoomi sekveneerimise lugemitest, testiti koostatud GM sündmuse-spetsiifiliste *k*-meeride nimekirju kolme simulatsiooni käigus saadud sekveneerimise toorlugemite andmestiku peal. Muundatud maisi sordi TC1507 puhul tuvastati kõik GM-spetsiifilised *k*-meerid, kui lugemite arv oli  $10^8$



(joonis 9). Antud lugemite arv võimaldab maisi genoomi seitsme kordset katmist. GM sordi MON87708 puhul tuvastati samuti kõik spetsiifilised  $k$ -meerid  $10^8$  lugemi juures (joonis 10), mis võimaldab soja genoomi 15 kordset katmist.



**Joonis 9.** Muundatud maisi sordi TC1507 kolme simulatsiooni puhul tuvastatud sordispetsiifiliste  $k$ -meeride arv erineva sekveneerimise lugemite hulga juures. Kokku on antud sordil 635 spetsiifilist  $k$ -meeri.

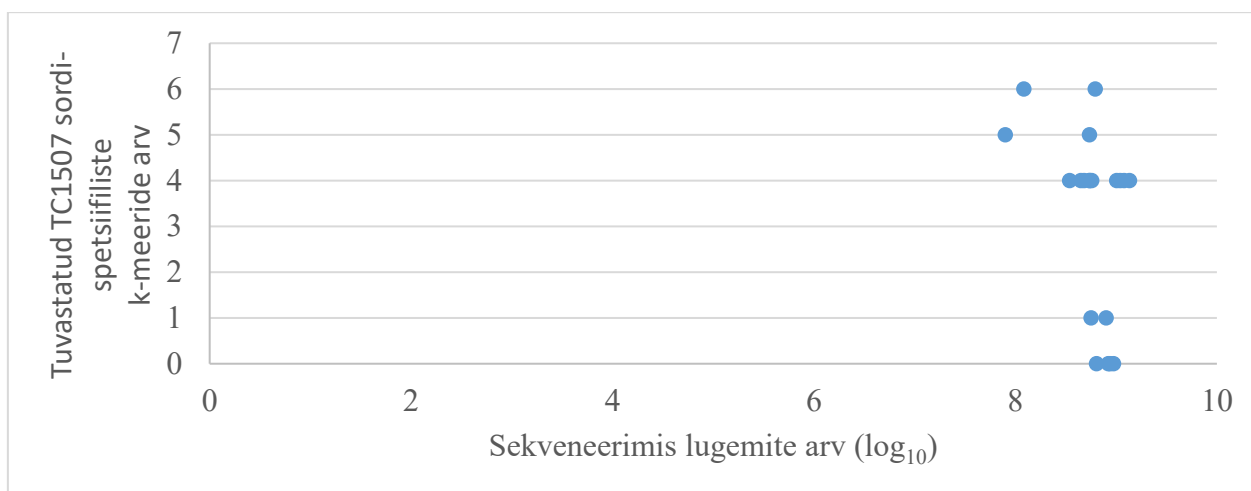


**Joonis 10.** Muundatud soja sordi MON87708 kolme simulatsiooni käigus tuvastatud sordispetsiifiliste  $k$ -meeride arv erineva sekveneerimise lugemite hulga juures. Kokku on antud sordil 1603 spetsiifilist  $k$ -meeri.

### 2.3.4 GM sordi-spetsiifilised $k$ -meerid muundamata genoomides

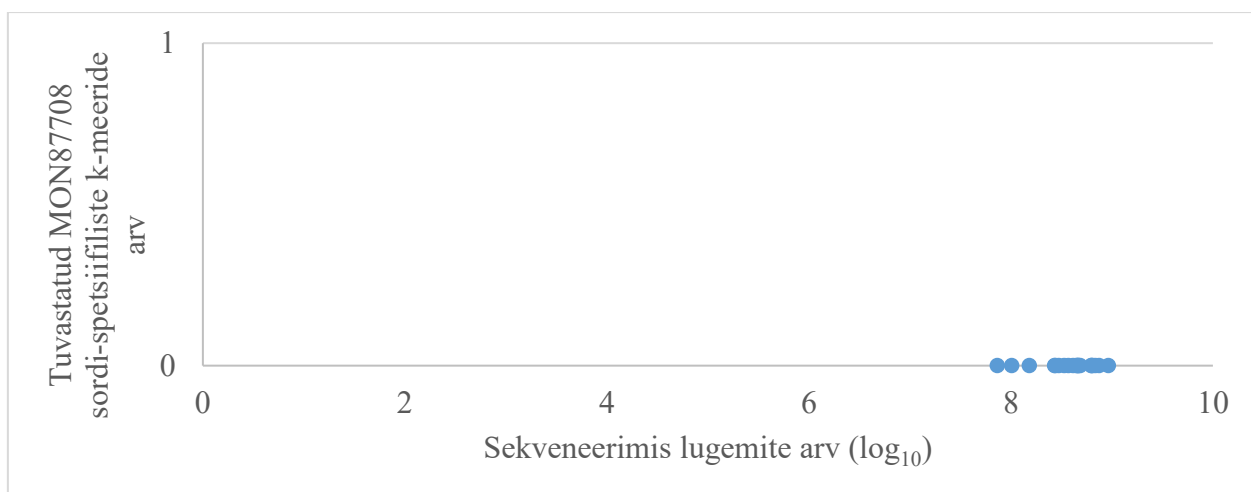
Loodud maisi ja soja GM sordi-spetsiifiliste  $k$ -meeride spetsiifilisust testiti kahekümne sama liigi muundamata täisgenoomi sekveneerimise toorlugemite abil, testimaks nende spetsiifilisust muundamise sündmustele.

GM maisi sordi TC1507 spetsiifilist 635  $k$ -meeri testiti 20 erineva maisi toorlugemi proovi peal. Leitud  $k$ -meeride arv varieerus nullist kuni kuueni proovi kohta (joonis 11).



**Joonis 11.** Maisi GM sordi TC1507 transgeeni-spetsiifiliste  $k$ -meeride ( $k = 32$ nt) esinemine muundamata maisi proovide toorlugemites.

GM soja sordi MON87708 sordi-spetsiifiliste  $k$ -meeride testimisel 20 erineval toorlugemi proovil. Nendes ühtegi antud GM sordi-spetsiifilist  $k$ -meeri ei tuvastatud (joonis 12).



**Joonis 12.** Soja GM sordi MON87708 transgeeni-spetsiifiliste  $k$ -meeride ( $k = 32$ ) esinemine muundamata maisi proovide toorlugemites.

### 2.3.5 Soja ja maisi liigispetsiifilised *k*-meerid

Soja kõigis kolmes kasutatud kloroplasti järjestuses esinevaid ühiseid 32-meere oli kokku 74 951. Pärast teiste liikide kloroplasti genoomi järjestustes esinevate *k*-meeride eemaldamist jäi alles 144 *k*-meeri ja pärast fülogeneetiliselt lähedaste sugulasliikide täisgenoomi sekveneerimise toorlugemites esinenud *k*-meeride eemaldamist jäi alles 126 soja spetsiifilist *k*-meeri. Järgnevalt testiti, kas need *k*-meerid on tuvastatavad kümne soja proovi täisgenoomi sekveneerimise toorlugemitest. Kõik 126 soja spetsiifilist *k*-meeri tuvastati kõikidest analüüsitud proovidest (Tabel 1).

**Tabel 1.** Soja liigispetsiifiliste *k*-meeride arv soja täisgenoomi sekveneerimise lugemites erinevate SRA andmebaasi proovide puhul.

NCBI SRA andmebaasi proov	Leitud <i>k</i> -meeride arv (126)
SRR1302613	126
SRR1302628	126
SRR2202261	126
SRR2424014	126
SRR5085797	126
SRR5479850	126
SRR5479862	126
SRR5479877	126
SRR5480106	126
SRR5480145	126

Maisi üheksast kloroplasti järjestusest vähemalt kaheksas esinevaid ühiseid 32-meere tuvastati 111 539. Sellest 74 100 *k*-meeri esinesid kaheksas ning 37 439 kõigis üheksas maisi kloroplasti järjestuses. Pärast teiste liikide kloroplasti genoomijärjestustes esinevate *k*-meeride eemaldamist jäi alles 4 805 *k*-meeri, milles 3 226 esinesid kaheksas ning 1 579 üheksas maisi kloroplasti järjestuses. Pärast fülogeneetiliselt lähedaste liikide sekveneerimisel saadud toorandmetes esinenud *k*-meeride eemaldamist jäi alles 1 503 *k*-meeri, millest 987 *k*-meeri esines kaheksas ning 516 üheksas maisi kloroplasti järjestuses.

Maisi spetsiifiliste *k*-meeride komplekti (kõigis üheksas maisi kloroplasti järjestuses esinevad *k*-meerid) testiti kümne SRA andmebaasist alla laetud maisi proovi täisgenoomi sekveerimise toorlugemitega. Kaheksal analüüsitud maisiproovil kümnest tuvastati 511 *k*-meeri 516-st maisi-spetsiifilistest *k*-meeridest, ühe proovi puhul 483 ja ühel proovil 448 *k*-meeri 516-st (tabel 2).

**Tabel 2** Maisi täisgenoomi sekveneerimise lugemitest tuvastatud liigispetsiifiliste *k*-meeride arv kümne maisi proovi puhul.

NCBI SRA andmebaasi maisi proov	Tuvastatud maisi-spetsiifiliste <i>k</i> -meeride arv (516)
SRR5976424	511
SRR7477706	511
SRR5976220	511
SRR065912	483
SRR7477683	511
SRR5725906	511
SRR065920	448
SRR5152975	511
SRR5152976	511
SRR5962677	511

### 2.3.6 GM sündmuse tuvastamine

Selleks, et tuvastada GM sündmust, juhul kui kättesaadaval on vähe informatsiooni, leiti GM sündmuse-spetsiifilised *k*-meerid (*k*-meerid, mis tekivad GM kasseti inserteerumisel taime genoomi ehk ei hõlma GM kasseti sees olevaid *k*-meere, aga identifitseerivad GM sündmuse). Antud nimekirjad loodi 28 sordile, millest 11 on GM soja ning 17 GM maisi sordid. Leitud sündmuse-spetsiifiliste *k*-meeride arv varieerus erinevate sortide vahel ja oli vahemikus 31-414 (tabel 3).

**Tabel 3.** GM maisi ja soja sündmuse-spetsiifilised  $k$ -meerid.

<b>GM sordi nimi</b>	<b>Sündmuse-spetsiifiliste <math>k</math>-meeride arv</b>
mais 31121	55
mais 32138	78
mais 4114	83
mais 59122	59
mais 98140	48
mais DAS40278	89
mais MIR162	91
mais MON87403	87
mais MON87411	80
mais MON87419	56
mais MON87427	129
mais MON87460	57
mais MON88017	85
mais MON89034	73
mais MZHG0JG	176
mais TC1507	123
mais VCO_O1981_5	52
soja CV127	57
soja DAS44406	54
soja DAS81419	83
soja DP356043	63
soja MON87701	74
soja MON87708	223
soja MON87712	55
soja MON87751	58
soja MON87769	61
soja MON89788	68
soja SYTHOH2	31

Kõiki loodud sündmuse-spetsiifilisi  $k$ -meeride komplekte testiti kasutades 20 muundamata maisi ja soja proovi täisgenoomi sekveneerimise toorlugemeid.

Maisi sündmuse-spetsiifiliste *k*-meeride nimekirjade testimisel ei leitud enamikest analüüsitud proovidest ühtegi GM päritolu *k*-meeri. Ülejäänud proovidest leiti kuni kaks GM spetsiifilist *k*-meeri. Üheksa sordi puhul leiti sündmuse-spetsiifilisi *k*-meere kolmest kuni viiest mitte GMO proovist. Viie sordi puhul leiti neid ainult ühest kuni kahest proovist. Kolme sordi *k*-meere ei leitud mitte ühestki toorlugemi proovist.

Soja sündmuse-spetsiifiliste *k*-meeride testimisel 18 sündmuse puhul ei leitud 20 proovist ühtegi *k*-meeri. Sündmuse CV127 puhul leiti ühest proovist 3 *k*-meeri ning sündmuse MON87712 puhul leiti ühest proovist üks *k*-meer .

## 2.4. ARUTELU

Üha suureneva muundatud taimesortide arvu tuvastamiseks arendatakse välja aina rohkem PCR-põhiseid detektsiooni meetodeid. Antud strateegia pole aga jätkusuutlik, et toime tulla kasvava muundatud sortide arvuga, seda puudust üritatakse ületada kasutades multiplex PCR, kuid selle pikaajalist edu on raske näha. Lisaks GM sortide arvu suurenemisele ennustatakse ka kasvupinna suurenemist, läbi biotehnoloogiliste põllukultuuride populariseerimisele arengumaades (ISAAA, 2017). Antud riikides puudub suure tõenäosusega ulatuslik GM testimise võimekus, mistõttu võib ennustada suurenenud GM kontaminatsiooni hulka Euroopa Liitu imporditavates põllusaadustes.

Antud töös töötati välja töövoog, mille abil saab tuvastada GM sordi-spetsiifilisi  $k$ -meere, mida saab kasutada selleks et hinnata GM sortidele omaseid  $k$ -meere. Esimene samm  $k$ -meeride põhise meetodi kasutamisel on valida optimaalne  $k$ -meeri pikkus. Mida pikem on kasutatava  $k$ -meeri pikkus, seda väiksem on antud  $k$ -meeride esinemisvõimalus mitte-sihtmärk järjestustes. Teisest küljest mõjutab pikemaid  $k$ -meere rohkem sekveneerimise platvormi ebatäpsused (Patro et al., 2014)  $k$ -meeride nimekirjade loomine ja töötlemine nõuab rohkem aega või arvutusvõimsust. Antud töös katsetati  $k$ -meeri pikkuseid 16 kuni 32 (joonis 5 ja joonis 6) ning otsustati kasutada 32-meere, kuna võimaldab leida enim sordi-spetsiifilisi  $k$ -meere.

GM sordi-spetsiifilised  $k$ -meeride nimekirjad koosnevad  $k$ -meeridest, mis ei esine muundamata doonor- ega peremeesorganismides ning üheski teises GM sordi järjestuses. GM sordi-spetsiifiliste  $k$ -meeride hulk (joonis 7 ja joonis 8) sordi kohta sõltub transgeense insertiooni pikkusest ja elementide modifitseerimise mahust võrreldes looduses oleva doonor järjestusega. Samuti sellest, kui palju erinevad kättesaadavate GM sortide insertiooni järjestused üksteistest. Kuna töö loomise hetkel ei olnud ligipääsu paljudele kommertsiaalses kasutuses olnud või olevatest GM põllukultuuride järjestustest, ei olnud võimalik täisautomatiseeritud meetodit GM sortide tuvastamiseks.

Töö käigus leitud soja ja maisi GM sordi-spetsiifiliste  $k$ -meeride testimiseks polnud kättesaadavaid GM taimede sekveneerimise andmeid, mistõttu tuli need kunstlikult töö käigus luua. Simuleeritud andmete analüüsi käigus hinnati ka, kui palju lugemeid on vaja kõikide GM sordi-spetsiifiliste  $k$ -meeride tuvastamiseks muundatud proovidest. Testitud tulemusel on näha, et kõikide GM sordi-spetsiifiliste  $k$ -meeride ülesleidmiseks on maisil vaja  $10^8$  lugemit (joonis 9) ning soja puhul  $10^8$  lugemit (joonis 10). Simuleerimisel on aga võrreldes päriseluga mitmeid puudusi, näiteks ei võta programm arvesse plastiidi genoomi varieeruvat koopiat arvu rakus. Ühes lehe rakus võib esineda 10 000 või isegi rohkem kloroplasti genoomi koopiat (Morley ja Nielsen, 2016) ning üle 100 mitokondri genoomi koopiat (Takanashi et al., 2010). Antud töös

saadud tulemus on pisut madalam kirjandusest leitavast, kus GM soja sündmuse toimumist suudeti tuvastada juba 5x genoomi katvuse juures (Lambirth et al., 2015). Et täpsemalt hinnata minimaalset vajalikku lugemite hulka, oleks vaja analüüsida rohkem proove, sh ka reaalseid GM taimede täisgenoomi sekveneerimisandmeid.

Selleks et hinnata maisi ja soja sordi-spetsiifiliste *k*-meeride spetsiifilisust, testiti loodud *k*-meeride nimekirju muundamata taime täisgenoomi sekveneerimislugemitega (joonis 12). GM maisi TC1507 sordi-spetsiifilisi *k*-meere leiti muundamata maisi toorlugemitest vaid vähesel määral (Joonis 11), GM soja MON87708 sordi-spetsiifilisi *k*-meere muundamata soja proovidest ei leitud. Arvestades valitud proovide lugemite arvu, mis soja puhul oli 13.7 miljardit lugemit ning maisi puhul 13.6 miljardit lugemit, on valepositiivsed tulemused muundamata maisi või soja taime proovis ebatõenäolised.

GM maisi ja soja tuvastamisel seguproovidest, nagu toidukaubad või sööt, tuleb täiendavalt kontrollida kas liik, antud juhul mais või soja, mille muundamisel GM sort loodi, esineb proovis. Liigi tuvastamisel saab kasutada kloroplastis asuvaid *k*-meere, mis võimaldab liike seguproovist tuvastada isegi väga väikese osakaalu puhul ning analüüsida uuritava GM seguproovi sekveneerimise toorlugemeid, et vältida genoomi töömahukat ja vigaderohket assambleerimist (Raime et al., 2020). Antud töö käigus leiti 126 maisi ning 516 soja liigispetsiifilist *k*-meeri. Tomati puhul on leitud 32-meeride arv 882 (Raime ja Remm, 2018). Taksonispetsiifiliste *k*-meeride arv sõltub fookuses oleva ja selle lähedaste liikide saadaval olevate assambleeritud kloroplasti genoomijärjestuste hulgast ja varieeruvusest. Leitud *k*-meere testiti 10 maisi (tabel 1) ja 10 soja (tabel 2) toorlugemi proovidel. Loodud liigispetsiifilist tuvastamist võimaldavad *k*-meeri nimekirjad vajavad päriselus rakendamiseks edasist testimist.

Sündmuse-spetsiifilised *k*-meerid võimaldavad sündmuse tuvastamist väheste informatsiooni juures, kuna esinevad ainult alas, mis on tekkinud GM kasseti insertioonil taime genoomi. *k*-meeride arv sündmuse-spetsiifiliste nimekirjades varieerub 31 kuni 223 *k*-meeri vahel (tabel 3). Antud *k*-meere leiti muundamata maisi ja soja sündmustest vähesel määral. GM kasseti integreerumisel taime genoomi tekib selle ümber „*filler DNA*“ lõik, mille maksimaalne pikkus on 51nt. Sellise pikkusega lõik suudab maksimaalselt hõlmata 82 erinevat 32-meeri. Selle järgi on teoreetiline maksimum 164 sündmuse-spetsiifilist *k*-meeri GM taimesordi kohta. See viitab vajadusele sügavamalt uurida muundamata sortide järjestusi ning transformeerimisel kasutatavaid plasmide. Vaatamata sellele saab peale katselist detektsiooni piiri määramist antud *k*-meere kasutada GM taimesortide tuvastamisel.

Käesoleva töö tulemusena on näha, et sordi- ja sündmuse-spetsiifilisi *k*-meere saab kasutada GMO tuvastamiseks. Kuid robustse meetodi väljatöötamiseks on vaja analüüsida oluliselt



rohkem GMO järjestusi. Nende kogumiseks ja haldamiseks tuleks luua andmebaas, mis sisaldaks GMO järjestusi, kus on selgelt defineeritud transgeensed kassetid ning ka erinevate elementide vahelised piirid, et võimaldada töös käsitlemata jäänud elementide tasandil tuvastamist. Seniks on GMO tuvastamiseks võimalik kasutada sündmuse-spetsiifilisi *k*-meere, mille leidmiseks ei ole vaja teiste GMO järjestuste olemasolu.

Selleks, et rakendada *k*-meeride põhiste lähenemist praegusel hetkel GMO tuvastamiseks tuleks, uuritavast proovist eraldatud DNA enne sekveneerimist rikastada, kasutades selleks „*DNA walking*“. Järgnevalt tuleks rikastatud proov sekveneerida ning toorlugemite andmetelt otsida GM *k*-meere. Rikastamine tõstab oluliselt GM päritolu *k*-meeride hulka proovis ning võimaldaks *k*-meeri põhiste rakendust kasutada juba praegu.

## KOKKUVÕTE

GMO rutiinseks tuvastamiseks kasutatakse enamasti PCR, mis on küll kättesaadav, kuid ei suuda oma tehniliste piirangute tõttu sammu pidada aina kiireneva GM sortide arendusega. Lahenduseks oleks üleminek sekveneerimise põhiste meetoditele. Siiani rakendatud joondamise põhised meetodid on aga oluliselt aeglasemad, kui  $k$ -meeri põhine lähenemine.

Käesolevas töös töötati välja töövoog GM sordi- ja sündmuse-spetsiifiliste  $k$ -meeride leidmiseks ning hinnati  $k$ -meeride kasutamise võimalikkust GM taimede tuvastamiseks. Esmalt kogudes internetis saadaolevad GM taimede järjestused ning hinnates nende erinevust üksteisest, keskendudes soja ja maisi muundatud sortidele. Järgnevalt loodi töö fookuses olnud liikide tuvastamist võimaldavad liigispetsiifilised  $k$ -meeri nimekirjad. Viimasena hinnati sündmuse-spetsiifiliste  $k$ -meeride arvu erinevates sortides, luues  $k$ -meeri nimekirjad, mida saab rakendada GM sündmuste tuvastamiseks.

GM taimede tuvastamiseks saab kasutada  $k$ -meere, kuid robustse tuvastusmeetodi loomiseks on vaja analüüsida oluliselt rohkem GM taimede järjestusi. Selle ületamiseks tuleks täiendada olemasolevaid GM organismide andmebaase.

# SUMMARY

## **Detection of modified organisms from raw sequencing data using *k*-mers of specific length**

Siimo Kangruoja

To increase crop yields further than traditional breeding methods enable, scientists have turned to genetic modifications. Legislation on the use and growth of such crops differs greatly between countries. That increases the risk of GM contamination in food and feed imports, especially in the EU, where the legislation process is more extensive and therefore slower.

PCR is routinely used for GMO detection, even though it is readily available, it cannot keep up with ever-increasing GM strain development. The solution would be a transition to sequencing-based methods. Until now only alignment-based methods have been implemented, which are slower than *k*-mer based methods.

In this work, the pipeline for the identification of GM specific *k*-mers was created and the feasibility of *k*-mer based GMO detection was tested. Firstly, by collecting all available GM crop sequences and evaluating their uniqueness, focusing on maize and soybean. This is important for having an overview of how distinguishable GM crop strains are from each other. Secondly, *k*-mer for species detection was created for maize and soybean. Lastly, the event-specific *k*-mer amount in different GM crop strains was created and evaluated.

Even though *k*-mer based detection enables a lot more than PCR based. For adoption in real life, a lot more data on GM sequences must be collected, because the absence of available data greatly hinders the creation of *k*-mer databases and by that the development of robust GMO detection methods. To overcome the limitation current databases must be greatly enhanced.

## **TÄNUAVALDUSED**

Täna enda juhendajat, Kairi Raimet, juhendamise eest käesoleva töö loomisel. Täna Mihkel Vaherit tehnilise nõu eest. Lisaks täna Maarja Lepametsa ja Lauris Kaplinski suupärase GenomeTesteri loomise eest.

## KASUTATUD KIRJANDUSE LOETELU

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.  
[https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Angers-Loustau, A., Petrillo, M., Bonfini, L., Gatto, F., Rosa, S., Patak, A., & Kreysa, J. (2014). JRC GMO-Matrix: a web application to support Genetically Modified Organisms detection strategies. *BMC Bioinformatics*, 15(1), 417.  
<https://doi.org/10.1186/s12859-014-0417-8>
- Arulandhu, A. J., van Dijk, J., Staats, M., Hagelaar, R., Voorhuijzen, M., Molenaar, B., ... Kok, E. (2018). NGS-based amplicon sequencing approach; towards a new era in GMO screening and detection. *Food Control*, 93, 201–210.  
<https://doi.org/10.1016/j.foodcont.2018.06.014>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Boyle, R. (2011). How To Genetically Modify a Seed, Step By Step. Retrieved from <http://www.popsoci.com/science/article/2011-01/life-cycle-genetically-modified-seed>
- Broeders, S., Papazova, N., den Bulcke, M. Van, & Roosens, N. (2012). Development of a Molecular Platform for GMO Detection in Food and Feed on the Basis of “Combinatory qPCR” Technology. In *Polymerase Chain Reaction* (pp. 363–404).  
<https://doi.org/10.5772/37898>
- Broeders, S. R. M., De Keersmaecker, S. C. J., & Roosens, N. H. C. (2012). How to Deal with the Upcoming Challenges in GMO Detection in Food and Feed. *Journal of Biomedicine and Biotechnology*, 2012, 1–11. <https://doi.org/10.1155/2012/402418>
- Carter, C. A., & Smith, A. (2007). Estimating the Market Effect of a Food Scare: The Case of Genetically Modified StarLink Corn. *The Review of Economics and Statistics*, (89(3)), 522–533.
- Chamberlain, J. S., Gibbs, R. A., Rainer, J. E., Nguyen, P. N., & Thomas, C. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Research*, 16(23), 11141–11156. <https://doi.org/10.1093/nar/16.23.11141>

- DIREKTIIV 2001/18/EÜ. (n.d.). *Geneetiliselt Muundatud Organismide Tahtliku Keskkonda Viimise Kohta Ja Nõukogu Direktiivi 90/220/EMÜ Kehtetuks Tunnistamise Kohta*.
- Dong, W., Yang, L., Shen, K., Kim, B., Kleter, G. A., Marvin, H. J. P., ... Zhang, D. (2008). *GMDD: a database of GMO detection methods*. 8, 1–8. <https://doi.org/10.1186/1471-2105-9-260>
- EUROPEAN COMMISSION. (2017). EU Register of authorised GMOs. Retrieved April 30, 2019, from [https://webgate.ec.europa.eu/dyna/gm\\_register/index\\_en.cfm](https://webgate.ec.europa.eu/dyna/gm_register/index_en.cfm)
- Fraiture, M.-A., Herman, P., De Loose, M., Debode, F., & Roosens, N. H. (2017). How Can We Better Detect Unauthorized GMOs in Food and Feed Chains? *Trends in Biotechnology*, 35(6), 508–517. <https://doi.org/10.1016/j.tibtech.2017.03.002>
- Fraiture, M.-A., Herman, P., Papazova, N., De Loose, M., Deforce, D., Ruttink, T., & Roosens, N. H. (2017). An integrated strategy combining DNA walking and NGS to detect GMOs. *Food Chemistry*, 232, 351–358. <https://doi.org/10.1016/j.foodchem.2017.03.067>
- Fraiture, M.-A., Herman, P., Taverniers, I., De Loose, M., Deforce, D., & Roosens, N. H. (2014). An innovative and integrated approach based on DNA walking to identify unauthorised GMOs. *Food Chemistry*, 147, 60–69. <https://doi.org/10.1016/j.foodchem.2013.09.112>
- Fraiture, M.-A., Herman, P., Taverniers, I., De Loose, M., Deforce, D., & Roosens, N. H. (2015). Current and New Approaches in GMO Detection: Challenges and Solutions. *BioMed Research International*, 2015(Table 1). <https://doi.org/10.1155/2015/392872>
- Fraiture, M.-A., Vandamme, J., Herman, P., & Roosens, N. H. C. (2018). Development and validation of an integrated DNA walking strategy to detect GMO expressing cry genes. *BMC Biotechnology*, 18(1), 40. <https://doi.org/10.1186/s12896-018-0446-x>
- Fraley, R. T., Rogers, S. G., Horsch, R. B., Sanders, P. R., Flick, J. S., Adams, S. P., ... Woo, S. C. (1983). Expression of bacterial genes in plant cells. *Proceedings of the National Academy of Sciences*, 80(15), 4803–4807.
- Gao, L., Cao, Y., Xia, Z., Jiang, G., Liu, G., Zhang, W., & Zhai, W. (2013). Do transgenesis and marker-assisted backcross breeding produce substantially equivalent plants? - A comparative study of transgenic and backcross rice carrying bacterial blight resistant gene Xa21. *BMC Genomics*, 14(1), 738. <https://doi.org/10.1186/1471-2164-14-738>
- GM Crop Events approved in United States of America. (n.d.). Retrieved May 10, 2019, from

[http://www.isaaa.org/gmapprovaldatabase/approvedeventsin/default.asp?CountryID=US  
&Country=United States of America](http://www.isaaa.org/gmapprovaldatabase/approvedeventsin/default.asp?CountryID=US&Country=United States of America)

- Grothaus, G., Bandla, M., Currier, T., Giroux, R., Jenkins, G., Lipp, M., ... Pantella, V. (2006). Immunoassay as an Analytical Tool in Agricultural Biotechnology. *JOURNAL OF AOAC INTERNATIONAL*, 89(4), 913–928.
- Holst-Jensen, A., Bertheau, Y., de Loose, M., Grohmann, L., Hamels, S., Hougs, L., ... Wulff, D. (2012). Detecting un-authorized genetically modified organisms (GMOs) and derived materials. *Biotechnology Advances*, 30(6), 1318–1335.  
<https://doi.org/10.1016/j.biotechadv.2012.01.024>
- Huang, W., Li, L., Myers, J. R., & Marth, G. T. (2012). ART: a next-generation sequencing read simulator. *Bioinformatics*, 28(4), 593–594.  
<https://doi.org/10.1093/bioinformatics/btr708>
- Hwang, H.-H., Yu, M., & Lai, E.-M. (2017). Agrobacterium -Mediated Plant Transformation: Biology and Applications. *The Arabidopsis Book*, 15, e0186.  
<https://doi.org/10.1199/tab.0186>
- ISAAA. (2017). *Global Status of Commercialized Biotech/GM Crops in 2017: Biotech Crop Adoption Surges as Economic Benefits Accumulate in 22 Years*. Ithaca, NY.
- Jackson, M. A., Sternes, P. R., Mudge, S. R., Graham, M. W., & Birch, R. G. (2014). Design rules for efficient transgene expression in plants. *Plant Biotechnology Journal*, 12(7), 925–933. <https://doi.org/10.1111/pbi.12197>
- James, C. (1997). *Global Status of Transgenic Crops in 1997*.
- James, C. (2015). ISAAA Brief 51-2015: Executive Summary. Retrieved September 28, 2016, from  
<http://www.isaaa.org/resources/publications/briefs/51/executivesummary/default.asp>
- Kamle, S., & Ali, S. (2013). Genetically modified crops: Detection strategies and biosafety issues. *Gene*, 522(2), 123–132. <https://doi.org/10.1016/j.gene.2013.03.107>
- Kaplinski, L., Lepamets, M., & Remm, M. (2015). GenomeTester4: a toolkit for performing basic set operations - union, intersection and complement on *k*-mer lists. *GigaScience*, 4(1), 58. <https://doi.org/10.1186/s13742-015-0097-y>
- Kim, S.-I., & Gelvin, S. B. (2007). Genome-wide analysis of Agrobacterium T-DNA integration sites in the Arabidopsis genome generated under non-selective conditions. *The Plant Journal*, 51(5), 779–791. <https://doi.org/10.1111/j.1365-313X.2007.03183.x>

- King, R. A., Read, D. S., Traugott, M., & Symondson, W. O. C. (2008). INVITED REVIEW: Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology*, 17(4), 947–963. <https://doi.org/10.1111/j.1365-294X.2007.03613.x>
- Košir, A. B., Arulandhu, A. J., Voorhuijzen, M. M., Xiao, H., Hagelaar, R., Staats, M., ... van Dijk, J. P. (2018). Author Correction: ALF: a strategy for identification of unauthorized GMOs in complex mixtures by a GW-NGS method and dedicated bioinformatics analysis. *Scientific Reports*, 8(1), 17645. <https://doi.org/10.1038/s41598-018-35950-y>
- Kregten, M. Van, Pater, S. De, Romeijn, R., Schendel, R. Van, Hooykaas, P. J. J., & Tijsterman, M. (2016). T-DNA integration in plants results from polymerase- $\theta$ -mediated DNA repair. *NATURE PLANTS*, 2(11), 1–6. <https://doi.org/10.1038/nplants.2016.164>
- Lambirth, K. C., Whaley, A. M., Schlueter, J. A., Bost, K. L., & Piller, K. J. (2015). CONTRAILS: A tool for rapid identification of transgene integration sites in complex, repetitive genomes using low-coverage paired-end sequencing. *Genomics Data*, 6, 175–181. <https://doi.org/10.1016/j.gdata.2015.09.001>
- Leinonen, R., Sugawara, H., & Shumway, M. (2011). The Sequence Read Archive. *Nucleic Acids Research*, 39(Database), D19–D21. <https://doi.org/10.1093/nar/gkq1019>
- Liang, C., van Dijk, J. P., Scholtens, I. M. J., Staats, M., Prins, T. W., Voorhuijzen, M. M., ... Kok, E. J. (2014). Detecting authorized and unauthorized genetically modified organisms containing vip3A by real-time PCR and next-generation sequencing. *Analytical and Bioanalytical Chemistry*, 406(11), 2603–2611. <https://doi.org/10.1007/s00216-014-7667-1>
- Lindgreen, S., Adair, K. L., & Gardner, P. P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Scientific Reports*, 6(1), 19233. <https://doi.org/10.1038/srep19233>
- Little, D. P. (2014). A DNA mini-barcode for land plants. *Molecular Ecology Resources*, 14(3), 437–446. <https://doi.org/10.1111/1755-0998.12194>
- Mano, J., Nishitsuji, Y., Kikuchi, Y., Fukudome, S., Hayashida, T., Kawakami, H., ... Kitta, K. (2017). Quantification of DNA fragmentation in processed foods using real-time PCR. *Food Chemistry*, 226, 149–155. <https://doi.org/10.1016/j.foodchem.2017.01.064>
- Markoulatos, P., Siafakas, N., & Moncany, M. (2002). Multiplex polymerase chain reaction: A practical approach. *Journal of Clinical Laboratory Analysis*, 16(1), 47–51.



<https://doi.org/10.1002/jcla.2058>

- Mazzara, M., Paoletti, C., & Corbisier, P. (2013). *Kernel Lot Distribution Assessment (KeLDA): a Comparative Study of Protein and DNA-Based Detection Methods for GMO Testing*. 210–220. <https://doi.org/10.1007/s12161-012-9407-5>
- Morley, S. A., & Nielsen, B. L. (2016). Chloroplast DNA Copy Number Changes during Plant Development in Organelle DNA Polymerase Mutants. *Frontiers in Plant Science*, 7. <https://doi.org/10.3389/fpls.2016.00057>
- NCBI Resource Coordinators. (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 45(D1), D12–D17. <https://doi.org/10.1093/nar/gkw1071>
- Novoa, E. M., Pavon-Eternod, M., Pan, T., & Ribas de Pouplana, L. (2012). A Role for tRNA Modifications in Genome Structure and Codon Usage. *Cell*, 149(1), 202–213. <https://doi.org/10.1016/j.cell.2012.01.050>
- Patro, R., Mount, S. M., & Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nature Biotechnology*, 32(5), 462–464. <https://doi.org/10.1038/nbt.2862>
- Perry, E. D., Ciliberto, F., Hennessy, D. A., Moschini, G., Qaim, M., Zilberman, D., ... Hennessy, D. A. (2016). Genetically engineered crops and pesticide use in U.S. maize and soybeans. *Science Advances*, 2(8), 900–902. <https://doi.org/10.1126/sciadv.1600850>
- Querci, M., Foti, N., Bogni, A., Kluga, L., Broll, H., & Van den Eede, G. (2009). Real-Time PCR-Based Ready-to-Use Multi-Target Analytical System for GMO Detection. *Food Analytical Methods*, 2(4), 325–336. <https://doi.org/10.1007/s12161-009-9093-0>
- Querci, M., Van den Bulcke, M., Žel, J., Van den Eede, G., & Broll, H. (2010). New approaches in GMO detection. *Analytical and Bioanalytical Chemistry*, 396(6), 1991–2002. <https://doi.org/10.1007/s00216-009-3237-3>
- Raime, K., Krjutškov, K., & Remm, M. (2020). Method for the Identification of Plant DNA in Food Using Alignment-Free Analysis of Sequencing Reads: A Case Study on Lupin. *Frontiers in Plant Science*, 11. <https://doi.org/10.3389/fpls.2020.00646>
- Raime, K., & Remm, M. (2018). Method for the Identification of Taxon-Specific *k*-mers from Chloroplast Genome: A Case Study on Tomato Plant (*Solanum lycopersicum*). *Frontiers in Plant Science*, 9. <https://doi.org/10.3389/fpls.2018.00006>
- Ripp, F., Krombholz, C., Liu, Y., Weber, M., Schäfer, A., Schmidt, B., ... Hankeln, T.

- (2014). All-Food-Seq (AFS): a quantifiable screen for species in biological samples by deep DNA sequencing. *BMC Genomics*, 15(1), 639. <https://doi.org/10.1186/1471-2164-15-639>
- Rizk, G., Lavenier, D., & Chikhi, R. (2013). DSK: *k*-mer counting with very low memory usage. *Bioinformatics*, 29(5), 652–653. <https://doi.org/10.1093/bioinformatics/btt020>
- Roosaare, M., Vaher, M., Kaplinski, L., Möls, M., Andreson, R., Lepamets, M., ... Remm, M. (2017). StrainSeeker: fast identification of bacterial strains from raw sequencing reads using user-provided guide trees. *PeerJ*, 5, e3353. <https://doi.org/10.7717/peerj.3353>
- Streatfield, S. J. (2007). Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnology Journal*, 5(1), 2–15. <https://doi.org/10.1111/j.1467-7652.2006.00216.x>
- Takanashi, H., Ohnishi, T., Mogi, M., Okamoto, T., Arimura, S., & Tsutsumi, N. (2010). Studies of mitochondrial morphology and DNA amount in the rice egg cell. *Current Genetics*, 56(1), 33–41. <https://doi.org/10.1007/s00294-009-0277-3>
- THE EUROPEAN COMMISSION. *Directive 2011/884/EU of COMMISSION IMPLEMENTING DECISION of 22 December 2011 on emergency measures regarding unauthorised genetically modified rice in rice products originating from China and repealing Decision 2008/289/EC.* , (2011).
- THE EUROPEAN COMMISSION. *COMMISSION STAFF WORKING DOCUMENT Genetically modified commodities in the EU Brussels, 8.3.2016 SWD.* , (2016).
- Thompson, J., Wailes, E., Durnd-Morat, A., & Leister, A. (2015). WELFARE EFFECTS OF U.S. LIBERTY LINK RICE CONTAMINATION. *Journal of Agricultural and Applied Economics*, 47(2), 243–259. <https://doi.org/10.1017/aae.2015.7>
- Vaher, M. (2016). *Bakteritüvede tuvastamine sekveneerimise toorlugemitest kindla pikkusega oligomeeride abil*. TARTU ÜLIKOOL.
- Weighardt, F. (2010). *The Analysis of Food Samples for the Presence of Genetically Modified Organisms Session 10 Quantitative PCR for the Detection of GMOs*.
- Wetterstrand, K. (2018). DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Retrieved May 1, 2019, from <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>
- Windels, P. (2003). T-DNA Integration in Arabidopsis Chromosomes. Presence and Origin of Filler DNA Sequences. *PLANT PHYSIOLOGY*, 133(4), 2061–2068.

<https://doi.org/10.1104/pp.103.027532>

- Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15(3), R46. <https://doi.org/10.1186/gb-2014-15-3-r46>
- Wright, D. A., Townsend, J. A., Winfrey, R. J., Irwin, P. A., Rajagopal, J., Lonosky, P. M., ... Voytas, D. F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *The Plant Journal*, 44(4), 693–705. <https://doi.org/10.1111/j.1365-313X.2005.02551.x>

## **KASUTATUD VEEBIAADRESSID**

[http://europa.eu/rapid/press-release\\_IP-19-161\\_en.htm](http://europa.eu/rapid/press-release_IP-19-161_en.htm) (külastatud: 25.07.2018)

<http://www.euginius.eu/> (külastatud: 18.03.2018)

<https://scholar.google.com/> (külastatud: 20.03.2018)

<https://www.lens.org> (külastatud: 20.03.2019)

<https://gmo-crl.jrc.ec.europa.eu/jrcgmoamplicons/> (külastatud: 20.03.2019)

<http://gmdd.sjtu.edu.cn/> (külastatud: 19.03.2018)

<https://www.ncbi.nlm.nih.gov/nuccore> (külastatud: 02.04.2018)

<https://www.ncbi.nlm.nih.gov/assembly> (külastatud: 02.04.2018)

# LISA

**LISA 1.** Antud töö käigus kasutatud GM sordi järjestused, mis võimaldavad sündmuse tasemel tuvastamist.

Sordi nimi(alternatiivne nimi)	Sündmuse unikaalne identifikaator
<b>Harilik Lutsern (<i>Medicago Sativa</i>)</b>	
J101	MON-ØØ1Ø1-8
J163	MON-ØØ163-7
KK179	MON-ØØ179-5
<b>Raps (<i>Brassica napus</i>)</b>	
61061	DP-Ø61Ø61-7
73496	DP-Ø73496-4
DHA	NS-B5ØØ27-4
GT73(RT73)	MON-ØØØ73-7
HCN28(T45)	ACS-BNØØ8-2
HCN92(Topas 19/2)	ACS-BNØØ7-1
MON88302	MON-883Ø2-9
MS1(B91-4)	ACS-BNØØ4-7
MS11(MS-B2)	BCS-BN012-7
MS8(MS-BN1)	ACS-BNØØ5-8
OXY-235	ACS-BNØ11-5
RF1 (B93-101)	ACS-BNØØ1-4
RF2(B94-2)	ACS-BNØØ2-5
RF3(RF-BN1)	ACS-BNØØ3-6
<b>Mehhiko puuvill (<i>Gossypium hirsutum L.</i>)</b>	
281-24-236	DAS-24236-5
3006-210-23	DAS-21Ø23-5
COT102(IR102)	SYN-IR1Ø2-7
COT67B(IR67B)	SYN-IR67B-1
GHB119	BCS-GHØØ5-8
GHB614	BCS-GHØØ2-5
GHB811(EE-GH7)	BCS-GH811-4
GK12	ei ole saadaval

<b>Sordi nimi(alternatiivne nimi)</b>	<b>Sündmuse identifikaator</b>	<b>unikaalne</b>
<b>Mehhiko puuvill (<i>Gossypium hirsutum L.</i>)</b>		
LLCotton25(EE-GH1)	ACS-GH001-3	
MON1445	MON-Ø1445-2	
MON15985	MON-15985-7	
MON531	MON-ØØ531-6	
MON757	MON-ØØ757-7	
MON88701	MON 887Ø1-3	
MON88702	MON-887Ø2-4	
MON88913	MON-88913-8	
T304-40	BCS-GHØØ4-7	
Xinmian 33B	ei ole saadaval	
GK-12	ei ole saadaval	
<b>Valge kastehein (<i>Agrostis stolonifera</i>)</b>		
ASR368	SMG-368ØØ-2	
<b>Baklažaan (<i>Solanum melongena</i>)</b>		
Bt Brinjal Event EE1	BARI Bt Begun-1, -2, -3 and -4	
<b>Harilik lina (<i>Linum usitatissimum L.</i>)</b>		
FP967 (CDC Triffid)	CDC-FLØØ1-2	
<b>Mais (<i>Zea mays L.</i>)</b>		
32138	DP-32138-1	
3272	SYN-E3272-5	
33121	DP-Ø33121-3	
4114	DP-ØØ4114-3	
5307	SYN-Ø53Ø7-1	
59122	DAS-59122-7	
98140	DP-Ø9814Ø-6	
Bt11 (X4334CBR, X4734CBR)	SYN-BTØ11-1	
Bt176 (176)	SYN-EV176-9	
BVLA430101	ei ole saadaval	
CBH-351	ACS-ZMØØ4-3	
DAS40278	DAS-4Ø278-9	

Sordi nimi(alternatiivne nimi)	Sündmuse unikaalne identifikaator
<b>Mais (<i>Zea mays</i> L.)</b>	
LY038	REN-00038-3
MIR162	SYN-IR162-4
MIR604	SYN-IR604-5
MON810	MON-00810-6
MON863	MON-00863-5
MON87403	MON87403-1
MON87411	MON-87411-9
MON87419	MON87419-8
MON87427	MON-87427-7
MON87460	MON-87460-4
MON88017	MON-88017-3
MON89034	MON-89034-3
MZHG0JG	SYN-000JG-2
MZIR098	SYN-00098-3
NK603	MON-00603-6
T25	ACS-ZM003-2
TC1507	DAS-01507-1
VCO-01981-5	VCO-01981-5
<b>Harilik papaya (<i>Carica papaya</i>)</b>	
55-1	CUH-CP551-8
<b>Riis (<i>Oryza sativa</i> L.)</b>	
bt63 (TT51-1 / Huahui-1)	ei ole saadaval
LLRICE601	BCS-OS003-7
LLRICE62	ACS-OS002-5
<b>Sojauba (<i>Glycine max</i> L.)</b>	
A2704-12	ACS-GM005-3
A5547-127	ACS-GM006-4
CV127	BPS-CV127-9
DAS44406-6	DAS-44406-6
DAS68416-4	DAS-68416-4
DP305423	DP-305423-1

<b>Sordi nimi(alternatiivne nimi)</b>	<b>Sündmuse identifikaator</b>	<b>unikaalne</b>
DAS81419	DAS-81419-2	
DP356043	DP-356043-5	
FG72 (FGØ72-2, FGØ72-3)	MST-FGØ72-3	
GTS 40-3-2 (40-3-2)	MON-Ø4Ø32-6	
MON87701	MON-877Ø1-2	
MON87705	MON-877Ø5-6	
MON87708	MON-877Ø8-9	
MON87712	MON-87712-4	
MON87751	MON-87751-7	
MON87769	MON87769-7	
MON89788	MON-89788-1	
SYHTØH2	SYN-ØØØH2-5	
<b>Punapeet (<i>Beta vulgaris</i>)</b>		
H7-1	KM-ØØØH71-4	
<b>Harilik Tomat (<i>Solanum lycopersicum</i>)</b>		
1345-4	ei ole saadaval	
Huafan No 1.	ei ole saadaval	



**LISA 2.** Kasutatud doonor- ja peremeesorganismide järjestused, koos allikaga.

<b>Liik</b>	<b>Andmebaas</b>	<b>Juurdepääsukood</b>
<i>Agrobacterium tumefaciens</i>	assembly	GCA_000971565.1
<i>Andropogon gerardii</i>	SRA	SRR10838875
<i>Arabidopsis thaliana</i>	assembly	GCA_000001735.2
<i>Arthrobacter globiformis</i>	assembly	GCA_000238915.2
<i>Avena sativa</i>	SRA	SRR9933148
<i>Bacillus licheniformis</i>	assembly	GCA_000011645.1
<i>Bacillus subtilis</i>	assembly	GCA_000009045.1
<i>Bacillus thuringiensis</i>	assembly	GCA_002119445.1
<i>Bacteriophage P1</i>	assembly	GCF_000844165.1
<i>Banana streak virus</i>	assembly	GCF_000867185.1
<i>Cassava vein mosaic virus</i>	assembly	GCF_000838625.1
<i>Cauliflower mosaic virus</i>	assembly	GCF_000848745.2
<i>Cestrum yellow leaf curling virus</i>	nucleotide	GCF_000845725.1
<i>Citrus yellow mosaic virus</i>	assembly	GCF_000838245.1
<i>Delftia acidovorans</i>	assembly	GCF_000018665.1
<i>Diabrotica virgifera virgifera</i>	assembly	GCF_003013835.1
<i>Discosoma sp</i>	SRA	SRR5131547
<i>Escherichia coli</i>	assembly	GCF_000008865.2
<i>Figwort mosaic virus</i>	assembly	GCF_000845905.1
<i>Glycine max</i>	assembly	GCF_000004515.5
<i>Gossypium barbadense</i>	assembly	GCA_008761655.1
<i>Helianthus annuus</i>	assembly	GCA_002127325.1
<i>Hordeum vulgare</i>	assembly	GCA_004114815.1
<i>Medicago truncatula</i>	assembly	GCF_000219495.3
<i>Neurospora crassa</i>	assembly	GCA_000786625.1
<i>Nicotiana tabacum</i>	assembly	GCF_000715135.1
<i>Oryza sativa</i>	assembly	GCA_009829375.1
<i>Peanut chlorotic streak virus</i>	assembly	GCF_000845345.1
<i>Petunia hybrida</i>	puudub	puudub
<i>Pisum sativum</i>	assembly	GCA_003013575.1

liik	andmebaas	juurdepääsukood
<i>Primula juliae</i>	puudub	puudub
<i>Saccharum officinarum</i>	SRA	SRR7771853
<i>Solanum tuberosum</i>	assembly	GCF_000226075.1
<i>Sorghum bicolor</i>	assembly	GCF_000003195.3
<i>Sphingobium herbicidovorans</i>	assembly	GCF_002080435.1
<i>Stenotrophomonas maltophilia</i>	assembly	GCF_000072485.1
<i>Streptomyces viridochromogenes</i>	assembly	GCF_000158955.1
<i>Tobacco etch virus</i>	assembly	GCF_000861345.1
<i>Tobacco mosaic virus</i>	assembly	GCF_000854365.1
<i>Triticum aestivum</i>	assembly	GCA_002220415.3
<i>Zea mays</i>	assembly	GCF_000005005.2

# LIHTLITSENTS

**Lihlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

**Mina, Siimo Kangruoja, (sünnikuupäev: 03.08.1993)**

1. annan Tartu Ülikoolile tasuta loa (lihlitsentsi) minu loodud teose „**Geneetiliselt muundatud taimede tuvastamine sekveneerimise toorlugemitest kasutades kindla pikkusega  $k$ -meere**“,

mille juhendaja on Kairi Raime,

- 1.1 reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Siimo Kangruoja*

*12.08.2020*